DOCUMENT RESUME

ED 126 142 95 TM 005 398

AUTHOR Hood, Paul D.; Blackwell, Laird
TITLE An Assessment System for Competence Based Education: The Educational Development, Dissemination, and Evaluation Training Program.
INSTITUTION Far West Lab. for Educational Research and Development, San Francisco, Calif.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C. Dissemination and Resources Group. School Practice and Service Div.
PUB DATE Dec 75
NOTE 314p.

EDRS PRICE MF-$0.83 HC-$16.73 Plus Postage.
DESCRIPTORS Certification; Decision Making; Diagnostic Tests; *Educational Development; *Educational Researchers; *Evaluation; Higher Education; Individualized Programs; *Information Dissemination; Instructional Materials; *Instructional Systems; Job Skills; Learning Activities; Manuals; Measurement Techniques; *Performance Based Education; Preservice Education; Professional Continuing Education; Professional Personnel; Professional Training; Research Skills; Self Evaluation; Simulation; Statistical Analysis; Test Construction; Tests; Test Validity; Work Experience
IDENTIFIERS *Far West DD&E Functional Competence Training Prog

ABSTRACT

This manual provides a description of the development and a guide to the use of the assessment resources developed in connection with the Far West Development, Dissemination, and Evaluation (DD&E) Functional Competence Training Program. The document concentrates on a user-oriented description of the content, validation, and use of the final version of the assessment system. DD&E is a competence-based professional development program providing a flexible array of training resources concentrating on entry level skills and knowledge required in DD&E. Chapter 1 describes the system, its design concepts, content, assessment methods, and overall use. Chapter 2 provides a detailed description of program planning assessment, including establishment of the student record file and the use of the DD&E Diagnostic Instrument. The development and validation data on this instrument is also summarized. Chapter 3 describes program progress assessment methods, discusses general guidelines for developing product ratings, and illustrates a student progress record matrix. Chapter 4 is concerned with existing competence certification. It describes and summarizes the field validation of a model Competence Assessment Battery that may be used to certify skills in "instructional product development". Three assessment methods are employed: ratings, knowledge tests, and job sample tests. Field test data indicate that the tests are valid in terms of content validity and in terms of ability to discriminate between subjects with and without actual product development/evaluation experience. (RC)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

# AN ASSESSMENT SYSTEM
# FOR COMPETENCE BASED EDUCATION:

## THE EDUCATIONAL DEVELOPMENT,
## DISSEMINATION, & EVALUATION
## TRAINING PROGRAM

Paul D. Hood - Laird Blackwell

December 1975

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
NATIONAL INSTITUTE OF EDUCATION
DISSEMINATION AND RESOURCES GROUP
SCHOOL PRACTICE AND SERVICE DIVISION

FAR WEST LABORATORY
FOR EDUCATIONAL RESEARCH AND DEVELOPMENT
1855 FOLSOM STREET • SAN FRANCISCO, CALIFORNIA

# AN ASSESSMENT SYSTEM FOR COMPETENCE BASED EDUCATION:

## THE EDUCATIONAL DEVELOPMENT, DISSEMINATION, & EVALUATION TRAINING PROGRAM

Paul D. Hood
Laird Blackwell

Produced by
FAR WEST LABORATORY FOR EDUCATIONAL
RESEARCH AND DEVELOPMENT
1855 Folsom Street
San Francisco, California 94103

Produced by the FAR WEST LABORATORY FOR EDUCATIONAL RESEARCH AND DEVELOPMENT, a public, nonprofit organization supported in part by funds from the Department of Health, Education, and Welfare, National Institute for Education. The opinions expressed in this document do not necessarily reflect the position or policy of the National Institute for Education, and no official endorsement by the Institute should be inferred.

Educational development is a new discipline. It involves, first, focusing on an important and specific area in need of improvement, and second, inventing, field testing, and validating a useful solution to that problem or need. The solution may be a self-contained product or process to be used by educators, by students, by parents, or by all of them together.

In 1970 the Far West Laboratory founded a consortium of R&D agencies and educational institutions. Our purpose was to develop and field test instructional models and materials and to train entry professional and paraprofessional personnel in educational development, dissemination, and evaluation (DD&E) competences. The original members of the FAR WEST CONSORTIUM FOR DD&E TRAINING included American Institutes for Research, Palo Alto; Educational Testing Service, Berkeley; Far West Laboratory for Educational Research and Development, San Francisco; Human Resources Research Organization, Monterey; Lockheed Missile and Space Agency, Educational Systems Division, Sunnyvale; and Stanford Research Institute, Palo Alto, as R&D agencies; and Monterey Peninsula College, Monterey; California State University, San Francisco; and San Mateo College District, San Mateo, as educational institutions. Between 1970 and 1973, the Far West Consortium for DD&E Training was funded by the Research Training Branch of the National Center for Research and Development, U.S. Office of Education. During this period, competence-based programs of DD&E training were initiated at both graduate and undergraduate levels.

In 1973 the National Institute for Education assumed sponsorship of the project, which was redirected from development and training to a development effort exclusively concerned with curriculum and assessment materials. The Far West Laboratory assumed primary responsibility, but continued to subcontract certain instructional module development with the American Institutes for Research and the Human Resources Research Organization. The final versions of these modules have been the responsibility of the Laboratory, but credit for their development is rightfully shared by all members of the Far West Consortium for DD&E Training. Each DD&E instructional module has completed several research and development cycles prior to release for reproduction and distribution. At least three phases of field testing--tryout of the prototype, a supervised field test, and an operational test under normal, user conditions without Laboratory participation--preceded formal external review and an official decision on acceptability.

The Laboratory's mission is to carry out surveys, research, development, and demonstrations in education, to disseminate information derived from such activities, and to provide services to schools and other educational agencies.

# FOREWARD

During the past few years, training institutions have become interested in "competence-based" or "performance-based" education. Currently, there is a major effort in teacher education and significant activity in other professional education areas. Beginning in 1970, the Far West Consortium for DD&E Training, a group of several educational R&D agencies and institutions of higher education located in the San Francisco Bay Area, undertook development and operation of a competence-based program for training educational development, dissemination and evaluation (DD&E) personnel. Initially, the Consortium focused on the development and implementation of undergraduate and graduate level programs that provided for joint planning, and curriculum development by faculty members of colleges and universities and by experienced DD&E staff of R&D agencies. When sponsorship of this project was transferred from the U.S. Office of Education to the National Institute of Education, the project emphasis shifted from curriculum development and operations to materials development and program diffusion. By the end of 1975, the project completed development and test of twenty-three instructional modules covering six content areas: planning, information/data collection and organization, communication, development, evaluation,dissemination and marketing. The instructional modules are augmented by several supporting documents which provide guidance for the planning, installation and operation of a variety of configurations of preservice and continuing education programs.

This document describes the DD&E Assessment System and the development and validation of three types of instruments (ratings, knowledge tests, performance simulation tests) which represent important components of the assessment system.

# ACKNOWLEDGEMENTS

The DD&E Assessment System described in this report evolved over several years and is the product of a number of persons.

From initial conceptualization of the project to its completion Dr. Susan Klein and Dr. John Egermeier of the National Institute of Education played significant roles in securing support and providing assistance. Their encouragement, patience and constructive criticism sustained our efforts throughout the duration of the project. Among the NIE appointed site visitors, Dr. John Hopkins, formerly of Indiana University and now at Research for Better Schools and Dr. Eva Baker, now Director of the Center for Study of Evaluation at UCLA, deserve special mention for the extraordinary quality of their evaluations and suggestions.

On the Far West Laboratory staff, many persons contributed. Dr. Joseph Ward, Dr. Wayne Rosenoff and Dr. John Hourigan developed and tested the "mini-batteries" which served as prototypes of the Competence Assessment Battery. Dr. Barbara Havassy and Ms. Andrea Lash developed and field tested the DD&E Diagnostic Instrument and also developed the preliminary field test version of the Competence Assessment Battery. Dr. Laird Blackwell supervised the field testing and data analysis, revised the Battery, developed the scoring manual and carried out all phases of the validation field testing, scoring, and analysis. He also contributed significantly to the preparation of this final report. Ms. Carol Burkhart, the project administrative assistant, maintained all project files, supervised the production of reports and instruments, and dealt with all of the technical aspects of mail-out and follow-up of the field tests. Ms. Lenni Ball supervised the production of this report.

We gratefully acknowledge the assistance of instructors and students of the following institutions who aided us in field testing one or more versions of the instrument in the DD&E Assessment System:

Brigham Young University
California State University, Long Beach
California State University, San Francisco
California State University, San Jose
Georgia State University
Michigan State University
Temple University
University of California, Berkeley
University of California, Los Angeles
University of Indiana

**

American Institutes for Research
Appalachia Educational Laboratory
Courseware, Inc.
Defense Language Institute
Far West Laboratory for Educational R&D
Human Resources Research Organization
Northwest Regional Educational Laboratory
Research for Better Schools
Teaching Research

In addition to the comments and critiques provided by field test subjects at the above institutions, we especially wish to acknowledge the assistance of the following persons who have reviewed and critiqued plans, instruments, or reports at various stages of this project:

Eva Baker, University of California, Los Angeles
Bela Banathy, Far West Laboratory
Richard Barbe, Georgia State University
Walter Borg, Utah State University
Leslie Briggs, Florida State University
Robert Gagne, Florida State University
John Hemphill, Far West Laboratory
Richard McCann, University of California, Los Angeles
Howard McFann, Human Resources Research Organization
Paul Madarasz, Defense Language Institute
David Merrill, Courseware, Inc.
Rodney Reed, University of California, Berkeley
Jack Sanders, Appalachia Educational Laboratory
Del Schalock, Teaching Research
Glenn Snelbecker, Temple University

To all who gave so much of their time and thought, my deepest appreciation.

P.D.H.

# EXECUTIVE SUMMARY

This manual provides a description of the development and a guide to the use of the assessment resources which have been developed in connection with the Far West Development, Dissemination and Evaluation (DD&E) Functional Competence Training Program.*

Previous reports have described the total DD&E program design (Hood, et al, 1970), the assessment system development plan (Hood and Banathy, 1973), and interim stages of progress (Hood, Havassy, Lash and Ward, 1973; Hood and Lash, 1974; Hood, Lash and Blackwell, 1974). This document concentrates primarily on a user-oriented description of the content, validation, and use of the final version of the assessment system.

The Far West DD&E Functional Competence Training Program is a competence-based professional development program which provides a flexible array of training resources which concentrate on "entry-level" skills and knowledge required in educational development, dissemination and evaluation (DD&E). One part of the program's effort focused on the development and validation of an assessment system which is designed to aid both student and instructor in: (1) program planning, (2) progress monitoring, and (3) exit assessment.

Chapter I describes the Far West DD&E assessment system, its design concepts, content, assessment methods, and overall use. One of the major design concepts is that the assessment system be decision-oriented, and that it be designed to facilitate the variety of decisions which must be made by students, instructors and employers. The assessment system provides a framework and the assessment resources for organizing a highly flexible, inexpensive and practical method for collecting and integrating information derived from a variety of sources. The three basic assessment methods are ratings, knowledge tests and job sample tests.

---

* Specifically excluded is any treatment of the assessment instruments which are directly associated with each of the 23 instructional modules.

Chapter II provides a detailed description of program planning assessment, including the establishment of the student record file and the use of the DD&E Diagnostic Instrument. This chapter summarizes the development and the validation data on this instrument, which are also described in detail in Appendix A The Diagnostic Instrument record of a student is examined in detail to illustrate its interpretation and use in student program planning.

Chapter III describes program progress assessment methods, briefly discusses general guidelines for developing product ratings, and also provides an illustration of a student progress record matrix.

Chapter IV is concerned with exit competence certification. It describes and summarizes the field validation of a model Competence Assessment Battery that may be used to certify skills in the area of "instructional product development." Three assessment methods are employed: ratings, knowledge tests and job sample tests. Field test data indicate that each of the instruments is valid, both in terms of content validity and in terms of ability to discriminate between subjects with and without actual product development/ evaluation experience.

Even where there are significant correlations between the three types of instruments, the relationships are very modest; hence each instrument tends to provide a different kind of valid information. Since many professional training programs seem to rely heavily on paper and pencil tests of knowledge, the results reported raise questions about the utility of assessment systems that rely exclusively or almost exclusively on such tests for measurement of knowledge. However, because the several different types of assessment instruments used in the Far West DD&E Assessment System are reasonably independent ofone another, and yet all are valid, their combined use should increase the overall validity of competence assessment decisions.

The several appendices to this document contain detailed information on the development, validation, and scoring of the individual instruments. The assessment model and the assessment resources described in this manual represent a moderately extensive but yet incomplete set of materials. We believe they are sufficient in scope and quality to provide a very good start in establishing a competence-based DD&E training program with practical assessment capacity; but there is a deliberate and clear challenge to make significant additions or modifications to the resources provided. For those not directly concerned with assessment of competencies in educational DD&E, the model and its validation data may provide ideas which can be adapted to assessment in other professional areas.

The manual has been written for three types of audiences: (1) persons with a general interest in assessment problems in competence-based education, (2) persons with specific interest in the assessment of R&D personnel, and (3) instructors who are adopting or adapting the Far West DD&E Functional Competence Training Program.

For those with general interest in competence-based educational assessment, we recommend Chapter I, Chapter II (omitting pp. 23-30), and Chapter III. This will provide a description of the rationale and method. Those with specific interest in the assessment of R&D personnel may wish to read Chapters I-IV, plus Appendices A and B, which describe the development and validation of the assessment instruments. Instructors who are adopting or adapting the Far West DD&E Functional Competence Training Program will also need to use the materials contained in the other appendices.

# TABLE OF CONTENTS

FOREWARD

ACKNOWLEDGEMENTS

EXECUTIVE SUMMARY

I. THE FAR WEST DD&E ASSESSMENT SYSTEM ........ 1

  A. Introduction ........ 1

  B. Scope and Focus of the Far West DD&E Competence Assessment System ........ 3

  C. Design Concepts ........ 5

  D. Assessment Methods ........ 9

  E. Overview of the Use of the Assessment System .... 11

II. PROGRAM PLANNING ........ 14

  A. Introduction ........ 14

  B. The DD&E Competence Matrix ........ 15

  C. The DD&E Diagnostic Instrument ........ 18

  D. Validation Test of the Diagnostic Instrument ... 23

  E. Interpretation and Use of the Diagnostic Instrument in Student Program Planning ........ 31

III. PROGRAM PROGRESS ASSESSMENT ........ 40

IV. EXIT COMPETENCE CERTIFICATION ........ 45

  A. Assessment of Exit Competence ........ 45

  B. Results of the Validation Test of the Battery .... 50

  C. Conclusions ........ 63

REFERENCES

APPENDICES

Appendix A: A Report on the Field Test and Revision of the DD&E Diagnostic Instrument

Appendix B: Development and Validation of DD&E (Development, Dissemination and Evaluation) Competence Assessment Battery

Appendix C: Instructional Product Development and Evaluation Competence Assessment Battery

Appendix D: Scoring Responses on the DD&E Knowledge Test

Appendix E: Scoring Manual for DD&E Simulation Tasks

Appendix F: Miscellaneous Forms

## LIST OF FIGURES

The Far West DD&E Assessment System

Figure 1 Assessment Instrument Development Priorities for the Functional Competence DD&E Training Program . . . . . . . . . . . . . . . . . . . . . 4

Figure 2 Derivation and Validation Linkages for a Catalogue of DD&E Competencies & Competence Assessment Instruments. . . . . . . . . . . . . . 6

Figure 3 DD&E Assessment Instrument Use. . . . . . . . . . 12

Program Planning

Figure 4 A Structural Matrix for the Organization of the DD&E Functional Context Training Program. . 17

Figure 5 DD&E Diagnostic Instrument Record Form. . . . . 20

Figure 6 Example of Use of DD&E Diagnostic Instrument Response Form As A Student Program Planning Device. . . . . . . . . . . . . . . . . . . . . . . . 35

Program Progress Assessment

Figure 7 Sample of Entries in A Student Progress Record Matrix . . . . . . . . . . . . . . . . . . . 43

Appendix A

Figure 1 DD&E Competence Matrix. . . . . . . . . . . . . . 2

## LIST OF TABLES

Program Planning

Table 1 DD&E Diagnostic Instrument Validation Subjects . 24

Table 2 Diagnostic Instrument Item Means and Significant Scores . . . . . . . . . . . . . . . . . . . . . . . 26

Table 3 DD&E Diagnostic Instrument - Means, Standard Deviations, Item Means, and Factor Analysis of Context, Process Skill, and Cluster Scales . . 29

Exit Competence Certification

Table 4 Biographic Data on the Validation Groups. . . . . 51

Table 5 RDD&E Activity Rating Scale Means and Significance (P-Level) of F-Tests for Validation Groups. . . . . . . . . . . . . . . . . . . . . . 53

Table 6 Means, F-Ratios and P-Levels for the F, P-F, and SC Total Scores on Job Knowledge Questions. . 56

Table 7 Means for Three R&E Experience Groups, F-Ratios and P-Levels for Tests of Differences Among the Groups on Simulation Test Scores. . . . . . . . . 60

Appendix A

Table 1 Cluster Analysis Scales . . . . . . . . . . . . . . . 9

Table 2 Diagnostic Instrument Item Means and Significance Tests. . . . . . . . . . . . . . . . 11

Table 3 DD&E Diagnostic Instrument - Means, Standard Deviations, Item Means, and Factor Analysis of Contest, Process Skill, and Cluster Scales. . . . 15

Table 4 Proficiency Levels for Students and For Developers by Subscales . . . . . . . . . . . . . . 19

Appendix B

Table 1 Biographic Data on the Validation Groups. . . . . 11

Table 2 RDD&E Activities Item Means . . . . . . . . . . . . 14

Table 3 RDD&E Activities Scale Intercorrelations, Means (M) Standard Deviations (SD) and Alpha Internal Consistency Coefficients (A). . . . . . . . . . . . 21

Table 4 Means, F-Ratios and P-Levels for the F, P-F, and SC Total Scores on Job Knowledge Questions. . . . 24

Table 5 Factor Loadings, Score Means and Standard Deviation (SD) for 27 Simulation Scores . . . . . 27

Table 6 Means for Three R&D Experience Groups, F-Rations and P-Levels for Tests of Differences on Simulation Test Scores. . . . . . . . . . . . . . . . 29

Table 7 Intercorrelations of Self-Ratings, Knowledge and Simulation Total Scales . . . . . . . . . . . . 31

# I. THE FAR WEST DD&E ASSESSMENT SYSTEM

## A. INTRODUCTION

This manual provides a description and guide to the use of assessment resources which have been developed in connection with the Development, Dissemination, and Evaluation (DD&E) Functional Competence Training Program. It also provides technical information on the development and validation of these assessment resources.

Competence-based education has become a "movement" that offers significant mechanisms for educational reform. Yet one of the most perplexing problems faced by those developing competence-based education programs is the definition and assessment of competence. In discussing Performance-Based Teacher Education (PBTE)* Elam observes:

> But the overriding problem before which the others pale to insignificance is that of the adequacy of measurement instruments and procedures. PBTE can only be successful if there are adequate means to assess the competency of the student. The bulk of the effort in establishing PBTE is most likely to go into the development of new instructional materials, into working out arrangements with the bursar and registrar, into devising ways for practicing teachers and administrators to share decision making, into moving the program into the field, and — most important of all — into developing ways to use faculty and librarians most effectively in the operation of unconventional [instructional] modules in a conventional system. *But when all this is done, an institution will still not have moved beyond current conventional grading procedures unless new methods are found for assessing the complex cognitive and affective objectives which are such an essential part of the training*...Yet this is the foundation stone on which the program rests...
>
> (Elam, 1971, pp. 21-22, emphasis added)

David Krathwohl, in an introductory note to Jack Merwin's paper (Merwin, 1973), concurs with these words:

> One can predict that performance-based teacher education (PBTE) is certain to fail to reach its ultimate objective if it continues its

---

*"Performance-based" and "competence-based" are often used synonomously. Some writers see "competence-based" as a more inclusive term and would reserve "performance-based" to a more restricted range of objectives and of abilities to perform. (Houston and Howsam, 1972.)

> present course. This failure will be caused by the almost complete lack of attention given to the assessment of teaching competencies, a core concept of PBTE. Only by such assessment can we achieve the goal of assuring that a teacher can indeed perform in ways that result in children learning.
>
> Until such assessment can be made, achieving PBTE will be a myth...
>
> These [assessment] problems require great concentration of thought, effort and research. Until and unless some real progress is made on resolving the problems of instrumentation and measurement, PBTE will go down in the history books as one more bandwagon in the long line of over-simplistic solutions for complex problems.
>
> (Merwin, 1973, pp. v, vi)

Although Elam and Krathwohl refer specifically to teacher education, their observations are equally pertinent for all competence-based or performance-based education. Specific details of the assessment problem may change, but not its general character.

The instructional and assessment resources developed by the Far West Consortium for Educational Development, Diffusion and Evaluation (DD&E) Training were designed to be employed in preservice training or continuing education of entry-level professionals. As Elam predicted, most of the Far West Consortium's effort went into the development of instructional materials, developing and evaluating new instructional systems, and the like. However, because we employed repeated cycles of development, testing and revision of DD&E instructional modules, the development of module tests was an early concern. Serious attempts to develop a comprehensive assessment system came at a much later stage in the project. We found the task to be immensely more difficult and costly than we first suspected. When confronted with the choice of preparing a comprehensive set of assessment instruments which we couldn't afford to validate or of rigorously developing and validating a much smaller set of instruments, we opted for validation of a "model" consisting of a core set of assessment items. We have attempted to provide guidance for the instructor who may (and should) want to prepare additional items following the model. The result is an "open" system. The assessment model and the assessment resources provided represent a moderately extensive but incomplete set of materials. We believe they are sufficient in scope and quality to provide a very good start in establishing a competence-based DD&E training program with a practical assessment capacity; but there is a clear challenge to make significant additions or modifications to the resources we have been able to provide.

## B. SCOPE AND FOCUS OF THE FAR WEST DD&E COMPETENCE ASSESSMENT SYSTEM

### 1. Definition of Competence

Currently, there seems to be no commonly-accepted approach to developing competence assessment instruments; in fact there is not even any commonly-accepted definition of "competence." Schalock and Thomas (1973) have made a useful distinction between two meanings commonly employed. One equates competence with the mastery of knowledges and skills assumed necessary to perform a particular function. The second holds competence to be the demonstrated ability to bring about outcomes specified in a given job description.

### 2. Criterion Levels

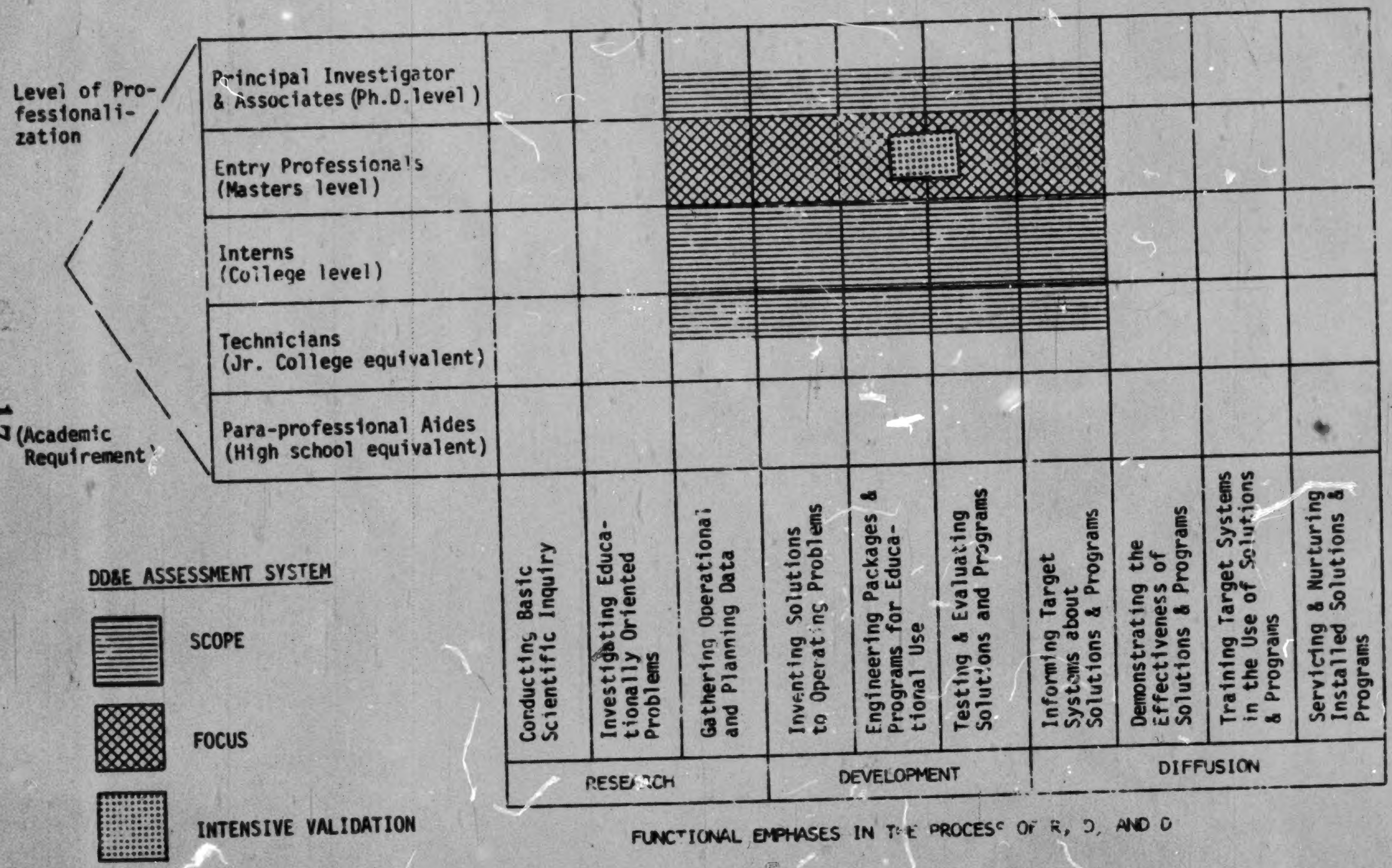
Richard Turner (1971) has provided a finer discrimination with six criterion levels. The six levels range from demonstration of mastery of knowledge and understanding (in level 6), demonstration of skill attainment in simple training and laboratory conditions (in levels 5 and 4), behaviors in actual conditions (in level 3), to evidence of short- and long-term pupil change (in levels 2 and 1). As we ascend Turner's criterion ladder, the problems and costs of assessment mount rapidly.

In the DD&E assessment system, we provide instruments at Turner's three lowest levels, namely at level 6, tests of knowledge, and at levels 5 and 4, performance of skills in job sample tests and simulations. Assessment at level 3, behavior in actual conditions, is provided only through ratings by supervisors. Thanks to recent work by Popham and others (1974), we may someday have instrumentation appropriate for measurement of short- and long-term effects of a developer's products on target audiences in levels 1 and 2. However, since test development and data gathering costs are so high at levels 1 and 2, we have not attempted assessment at these criterion levels.

It may be helpful to look at the competence content area addressed by the Far West project. (See Figure 1, adapted from Clark and Hopkins, 1969, p.14.)

Figure 1

ASSESSMENT INSTRUMENT DEVELOPMENT PRIORITIES FOR THE FUNCTIONAL COMPETENCE DD&E TRAINING PROGRAM



| Level of Professionalization (Academic Requirement) | Conducting Basic Scientific Inquiry | Investigating Educationally Oriented Problems | Gathering Operational and Planning Data | Inventing Solutions to Operating Problems | Engineering Packages & Programs for Educational Use | Testing & Evaluating Solutions and Programs | Informing Target Systems about Solutions & Programs | Demonstrating the Effectiveness of Solutions & Programs | Training Target Systems in the Use of Solutions & Programs | Servicing & Nurturing Installed Solutions & Programs |
|---|---|---|---|---|---|---|---|---|---|---|
| | RESEARCH | | | DEVELOPMENT | | | DIFFUSION | | | |
| Principal Investigator & Associates (Ph.D. level) | | | ▤ | ▤ | ▤ | ▤ | ▤ | | | |
| Entry Professionals (Masters level) | | | ▦ | ▦ | ▦ ░ | ░ ▦ | ▦ | | | |
| Interns (College level) | | | ▤ | ▤ | ▤ | ▤ | ▤ | | | |
| Technicians (Jr. College equivalent) | | | ▤ | ▤ | ▤ | ▤ | ▤ | | | |
| Para-professional Aides (High school equivalent) | | | | | | | | | | |

FUNCTIONAL EMPHASES IN THE PROCESS OF R, D, AND D

DD&E ASSESSMENT SYSTEM

- ▤ SCOPE
- ▦ FOCUS
- ░ INTENSIVE VALIDATION

Figure 1 locates the scope, focus, and area of concentration. Our focus is on competencies required of entry-professionals (masters degree level personnel) in the area of development, with some spill-over into immediately adjacent areas of research and diffusion. The reason for this focus is that it corresponds to the area addressed by the training materials we have developed. Figure 1 indicates that the scope — the area where the assessment materials may be useful — extends to a wider range of professionalization levels. Simply stated, although the competence criterion levels focus on competence appropriate for entry-level professionals, the assessment instruments may provide useful information about levels of competence which are below or slightly above those considered appropriate for DD&E entry-level professionals. Finally, Figure 1 indicates that we made a concerted effort at intensive instrument development and validation in the more restricted sub-areas of "engineering packages and programs for educational use" and "testing and evaluating solutions and programs."

Figure 2 indicates that the project, in its attempt to derive curriculum and instructional objectives, has drawn on three data bases, the AERA Task Force on Training (Worthen, et al., 1971, 1973), the Oregon Studies in Educational RDD&E (Schalock et at., 1973) and the Far West Consortium's task analyses (Hood, et al., 1970). Parallel development and field testing of instructional modules and of test items provided field test data on students as well as expert review of content. This experience was fed back in revision cycles resulting in modification of both the instructional resources and the assessment instrumentation.

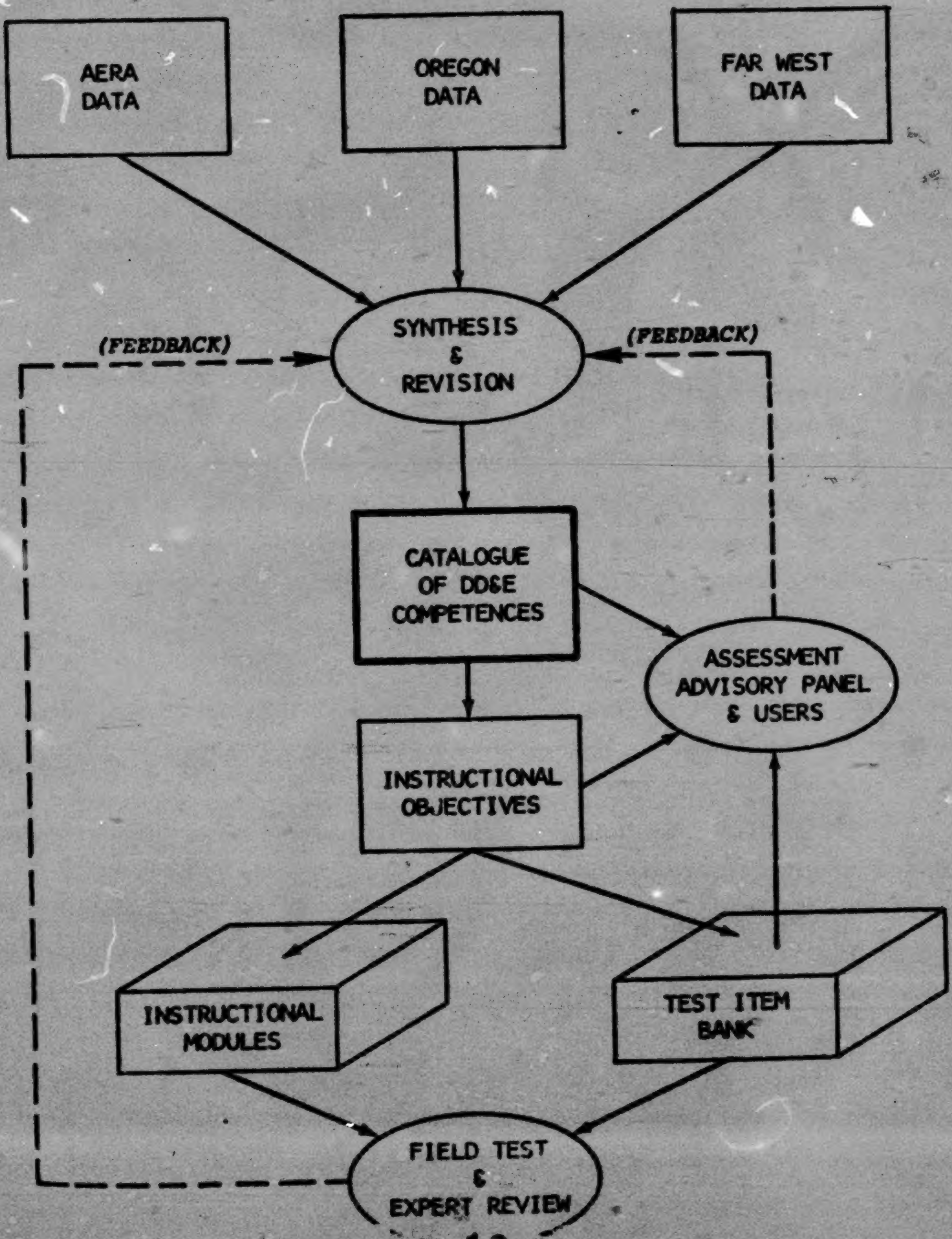
## C. DESIGN CONCEPTS

One of the major design concepts in our approach is that the assessment system be decision-oriented. We are not so much concerned about tests and their psychometric properties as we are about the utility and feasibility of competence-based decision making. Who are the decision makers in a competence-based educational system?

First, there are the students. They are concerned with questions regarding status and progress. They want to know what competencies they have already mastered; what their level of proficiency is vis a vis training exit or employ-

Figure 2

Derivation and Validation Linkages for a Catalogue of DD&E Competencies & Competence Assessment Instruments.


AERA DATA
OREGON DATA
FAR WEST DATA
(FEEDBACK)
SYNTHESIS & REVISION
(FEEDBACK)
CATALOGUE OF DD&E COMPETENCES
ASSESSMENT ADVISORY PANEL & USERS
INSTRUCTIONAL OBJECTIVES
INSTRUCTIONAL MODULES
TEST ITEM BANK
FIELD TEST & EXPERT REVIEW

ment requirements; and whether they should repeat a unit of instruction or proceed to the next unit.

Second, there are the instructors. They want to know the students' current levels of accomplishment and skills; what further training should be planned; what progress is being made; and whether in fact exit mastery is attained.

A third group are the employers. They are particularly interested in the match between a prospective employee's attainments and the job requirements. Does the person qualify for a particular job or will further on-the-job training be needed? Given valid and detailed information, employers may restructure the work for more effective use of their employees' current levels of competence or possibly may encourage more on-the-job training or inservice education.

Finally, there are those of us who are concerned with the development and evaluation of competence-based training programs and resources. They need to know about the range of entry-level knowledges, skills, and sensitivities. Do specific instructional materials and methods facilitate attainment of particular competencies? How can instructional materials and methods be improved? How can credible evidence of program performance, sufficiently impressive to persuade potential users to adopt the program or to hire its graduates, be provided?

These are some of the questions that students, instructors, employers and developers may ask. How are the answers to be provided?

We have already indicated that one of our major design concepts is that the assessment system be decision oriented. This in turn implies that we be concerned with utility, usability, validity, and reliability; and in that order. If the instrumentation does not lead to better decision making, for all the decision makers, there is little chance that the assessment system will actually be implemented and even less chance that it will be maintained. So, above all, the system must be useful. It must have apparent and real utility in helping users make better decisions. The usability of a system means that time, costs, psychological threat, etc., must be minimized while maximizing the information obtained for each stakeholder — student, instructor, potential or actual employer. In our opinion, the requirements for utility and usability transcend those for validity and reliability. Obviously, there must be some non-trivial level of validity and reliability. But if one takes a decision-theoretic approach, it is the utility of the decision that is paramount. The major issues are the risks of error and the costs of those errors. The challenge for us has been how to configure a set of information gathering instruments and processes in a form that would lead to their acceptance and use and also provide for efficient organization of information that would have an actual bearing on decisions.

Three important and highly practical, technical points are that: 1) we are dealing with a classification model; 2) we are dealing with a multiple-sequential decision model; and 3) a multiple-sequential decision model may be best handled with a Bayesian statistical decision approach.

Many of our assessment approaches are based on selection models where we attempt on the basis of a one-time assessment to decide whether a student should be selected into or out of a training program, passed in a course or certified for graduation. Although selection remains an aspect of competence-based programs, the assessment problems really deal more with classification. The instructor, the student and the employer have different perspectives and the majority of their decisions are not simple "go, no-go" decisions. More often, it is a matter of deciding how well prepared, which job, what kind of a career, how much and what kind of supervision, what potential for advancement, what kind of instructional resources, etc.

The competence-based, individualized approach provides the need and the opportunity to make a sequence of tests and decisions. Few of them are totally irreversible. This is a fortunate situation since few of the measures which are feasible provide highly reliable or highly valid measures when used singly or on one occasion. We know from decades of personnel research work that interrater reliabilities of complex performances and products may not go much above .6 or .7 in actual practice. And validity correlations of a specific predictor to a specific criterion are often in the .2 to .4 range. While this situation is tolerable when dealing with groups of persons, it becomes less so when dealing with decisions specific to one person. However, if we take a relatively large number of measures over time and employ a variety of methods, the cumulative sequential decision process itself can attain a much higher reliability and validity with correspondingly lower classification errors and costs.

We have known for some time that, at least in theory, the Bayesian statistical approach was an attractive alternative to the classical approach, since it is rarely the case that any decision maker has a flat prior expectation with total ignorance regarding the probabilities of alternative outcomes. Recent contributions by Brown (1969), Ferguson and Novick (1973) Swaminathan, Hambleton, Algina (1975), and others, suggest that eventually we shall be able to provide relatively simple procedures for all decision makers, whether they are students, instructors, or employers, to arrive at more effective decisions through fuller use of available information.

For the above reasons, we chose to focus on the idea of developing a highly flexible, inexpensive and practical assessment system rather than a simple collection of test instruments. Indeed, this perspective caused us to be far more concerned about helping decision makers to use information more effectively than in simply developing more reliable and valid measurement devices. Please understand that we are not underrating the need for better measurement, but simply asserting that from a very practical point of view, the crucial problem in competence-based education is to get people, including instructors and students, to want and know how to use assessment information. Until this happens, we may have technically impeccable but unusable competence assessment instrumentation. Hence, the objective was to make the procedures attractive, inexpensive, easy to use, easy to understand, face valid, useful and meaningful.

## D. ASSESSMENT METHODS

### 1. Three Major Methods

After review of a number of alternatives, we selected three basic assessment methods: ratings, knowledge tests, and job sample tests.* Although these methods can be augmented by biographical information, questionnaires, interviews, and other sources of information, these three methods constitute the foundation for assessment.

---

*Initially we also intended to develop some general form for rating any DD&E "products". However, experience based on three separate field tests of several versions of simulation tasks, all requiring test subjects to produce various "products", has thoroughly discouraged us in this particular venture.

## 2. Ratings

We have found that while ratings by students, instructors, peers, and work supervisors are useful; possibly only the student rating may be feasible in the preservice program. Although subject to well-known types of error, extensive experience in the use of ratings indicates that if well-designed and properly employed they can yield modestly reliable and valid information at very low cost. After trying several formats we selected a descriptive, performance-based rating scale which permits relatively unambiguous determination of observable or easily inferable levels of performance.

## 3. Job Knowledge Tests

Initially, we attempted to develop highly objective and easy-to-score job knowledge tests. However, we found it difficult to write items testing for comprehension, ability to make applications or evaluative judgments with completely objective formats. Our selected approach is to use essay and short-answer written questions, which call for more complex constructed responses on the part of the testee and also for greater thought and attention on the part of the grader. This is one area where we have accepted increased grading costs in the hope of obtaining greater meaningfulness and utility.

## 4. Job Sample Tests

Ratings and job knowledge tests are relatively easy to develop and can cover a lot of "competence territory" in relatively short time. Job samples are something else. Generally, they are quite time consuming and expensive to administer and to score. Moreover, it is quite difficult to avoid a certain degree of situational specificity which may significantly limit the generality of results. After considerable experimenting, we developed a simulation test which consists of a series of separately scorable but logically related job samples.

## E. OVERVIEW OF THE USE OF THE ASSESSMENT SYSTEM

Our concerns for costs and practical use, led to a sequential strategy in which the least expensive devices such as self-ratings are used most extensively, but with cross checking against more expensive sources of information such as knowledge tests and job samples. An item sampling strategy coupled with a Bayesian decision approach in which test items are selected for their relative potential in reducing the decision makers uncertainty is employed.

The various types of instruments and their use are suggested in Figure 3, Please note that we have created an artificial dichotomy to illustrate a difference between diagnostic use and assessment use. In the earlier stages of the DD&E Training Program, the emphasis is primarily on program planning, guidance, and counseling. As the student progresses, the emphasis tends to shift to concerns about completion of modules, attainment of objectives, credit for attainment, and ultimately to graduation and competence certification.

Comparable data, and sometimes identical or parallel test items and instruments, may be employed for both uses. A student's file is "opened" by recording pertinent biographical, academic, and work experience data from the student's application form. This may be augmented by instructor interview data. The first structured instrument the student encounters is the Diagnostic Instrument of 72 self-rating items which has been designed to reveal a competence profile on the three DD&E functional contexts (development, dissemination, and evaluation) and on six process skills (analyzing, planning, producing, evaluating, collecting and organizing information, and communicating). On the basis of this information taken together (remember we are employing a sequential, Bayesian, item sampling strategy), the instructor may decide to probe areas where further information is needed — perhaps, because the students' self appraisals in those areas suggest discrepancies in either being higher or lower than expected. The least expensive alternatives available are: 1) interview, 2) use of a second tailored self rating drawn from the DD&E item pool or the use of one or more of the DD&E module tests. In some instances, supervisors' ratings, job samples or products may be available. However, their use in diagnostic situations would be unusual except in continuing education programs or in programs providing simultaneous intern or on-the-job training.

FIGURE 3

DD&E ASSESSMENT INSTRUMENT USE

| INSTRUMENT TYPE | DIAGNOSTIC USE<br>• Guidance<br>• Counseling<br>• Program Planning | ASSESSMENT USE<br>• Progress Assessment<br>• Credit-by-examination<br>• Certification of Mastery |
|---|---|---|
| 1. Biographic Data | DD&E BIOGRAPHIC DATA FORM | (Same biographic data) |
| 2. Interview Data | Instructor's initial program planning interview | Instructor's interview file |
| 3. Academic Data | Student's application form; admission record | Program progress information |
| 4. Work Data | DD&E BIOGRAPHIC DATA FORM | Student intern or work application data if applicable |
| 5. Ratings | DD&E DIAGNOSTIC INSTRUMENT | |
| a. Self ratings | (DD&E ITEM POOL) | DD&E ITEM POOL |
| b. Supervisor ratings | (Use supervisor scale with Diagnostic Instrument if supervisor rating can be obtained) | DD&E ITEM POOL (supervisor ratings if applicable) |
| c. Instructor ratings | N/A | DD&E ITEM POOL |
| 6. Knowledge Tests | DD&E MODULE TESTS may be used as pre-tests if indicated | DD&E MODULE TESTS<br>DEVELOPMENT BATTERY KNOWLEDGE TEST<br>Instructors Tests |
| 7. Product Ratings | Instructor appraisal of DD&E products or reports student may have produced | DEVELOPMENT SIMULATION TASKS<br>Instructor's ratings of student's products |

As students progress through the program they will encounter up to twenty-three end-of-module knowledge tests. The students may also produce a variety of rateable products as part of application projects. In some instances, they may have an internship where a valid job supervisor rating can be obtained.

At the end of the program this cumulative file of information can be augmented by a final self-rating. Because of the flexible item sampling format, this final self-rating may be tailored to probe areas corresponding to the student's own program objectives. When relevant, supervisor or instructor ratings may also be obtained. Finally, for those students who have focused on development and evaluation, we have provided a more comprehensive job knowledge test and job sample (simulation) test.

Validation was approached from two directions. First a panel of experts, including DD&E work supervisors, made judgments about face falidity and relevence. Second, we required that the measured differences between groups known to have markedly different competence levels be significant and of practically meaningful magnitudes. The technical appendices to this guide report the results of our validation studies.

# II. PROGRAM PLANNING

## A. INTRODUCTION

A fundamental assumption of the DD&E Functional Competence Training Program, no matter what its operational configuration (see the DD&E Program Instructional Guide) might be, is that the Program will be characterized by some reasonable effort to provide an individualized program of instruction tailored to the particular needs and interests of each student. Moreover, the Program deliberately modifies the role of the instructor from that of a "presenter " of information to that of a "mentor" who counsels and guides students in planning their program of study, assists them in selecting and using instructional resources, and aids them in assessing and evaluating attained competence.

This is a role which calls for use of vastly more information about the status and progress of each student than is the case in more traditional instructional systems. Given the realities of instructor time and effort, there are obviously severe limits in terms of what the instructor can reasonably accomplish in developing, maintaining and using such information. We have tried to keep this in mind in designing the assessment system.

Our first advice is to create a separate file for each student. The file may eventually contain the following items:

- a summary status sheet on work accomplished
- a summary status sheet on estimated competence levels
- instructor's notes (initial interview, subsequent interviews, comments, appraisals, etc.)
- the program admissions records (application, transcripts, test results, letters, etc., depending on the requirements of the institution and department)
- copy of the Diagnostic Instrument record
- DD&E Training Program Biographic Data form
- end-of-module tests, job knowledge tests, simulation tests, and other assessment records
- copies of reports, papers, project descriptions, etc. documenting student work
- letters, ratings, and other information supplied by intern supervisors or employers if the program provides an application component

Because such a file can become quite voluminous, the two status sheets are of particular importance since they can provide, at a glance, an overview of where the student is and what activities have been accomplished. Our assumption is that the instructor will refer to this file as needed in connection with each program planning or program review session and that the instructor will make at least brief notations in the instructor's notes section on a frequent basis and certainly in connection with every planning or review session.

Appendix F contains model copies of several forms. Because we have envisioned a flexible program in which the Far West DD&E materials would be used in conjunction with other resources, it must be emphasized again that all the assessment forms and instruments which are provided should be treated as models or outlines that should be modified, and in most cases expanded, to reflect the specific characteristics of each institution's instructional program.

## B. THE DD&E COMPETENCE MATRIX

In assessing current status or planning for further study, a basic problem for both the student and the instructor is how to view the domain of DD&E competence. The AERA, Oregon Studies, and Far West data all indicate that the DD&E domain is complex and consists of a number of different areas. However, there is evidence that one can discriminate between "development," "dissemination," and "evaluation," at least in the sense that one or more factors can be relatively unambiguously associated with each of these commonly-recognized functional areas. The Far West data would suggest that, at least in terms of self ratings of levels of competence, evaluation is the most homogeneous area and development is the most heterogeneous. Stated more exactly, most of the self-rating items referring to evaluation tasks load on only one or two factors while items describing development load on several different factors.

The original design for the DD&E Program (Hood, Banathy, Ward, et al., 1970) assumed that the contexts of DD&E would be significant. However, we also assumed that competence in all three DD&E contexts would depend on more generic skills, such as collecting and organizing information or communicating, that would be essential for effective performance in any context. Consequently, we developed a structural matrix for the organization of the DD&E Functional Context Training Program which consisted of both contexts and skills. (See Figure 4.)

Various versions of this matrix have had different numbers of, and labels for, rows and columns as we searched for appropriate categories to define context and skills. The DD&E instructional resources have been organized, and interrelated, to reflect both the context and the skills approaches, consequently, we attempted to develop an assessment system that would be compatible with this dual orientation.

Because the initial focus of assessment efforts will be on Program Planning, the instructor must have a good grasp of the overall instructional program.

The Instructional System and its rationale are described in other DD&E documents, but it may be helpful to review a bit of its structure at this point. There is a total of 23 instructional modules organized in six major series, three series are primarily context oriented and three are primarily skills oriented. Although the modules have been deliberately designed to afford considerable flexibility regarding the number and sequence in which students may use the modules in their program of study, there is a general plan.

The Planning Series provides an orientation for the entire DD&E Program, since it entails consideration of all the phases and contexts of a development project. The modules in this series focus on the earlier phases of development (e.g., specification of outcomes and consideration of alternatives) and also (optionally) provide introductions to the three major context series:

- Development
- Field Test and Evaluation
- Dissemination and Marketing

Most of the (generic, i.e., cross-context) skills (e.g., analyzing, planning, producing) are treated within the contexts of each of these three series. However, in addition to planning skills, two other generic skills areas were selected for intensive treatment: collection and organization of information

### Figure 4

### A Structural Matrix for the Organization of the DD&E Functional Context Training Program

| CONTEXTS / SKILLS | Development | Field Test and Evaluation | Dissemination and Marketing |
|---|---|---|---|
| Collecting Information | | | |
| Analyzing | | | |
| Planning | | | |
| Producing | | | |
| Evaluating | | | |
| Communicating | | | |

and data, and communication (oral and written including preparation of reports and instructional writing).*

It should be noted that although the organization of the instructional series has singled out three generic skills areas for special independent treatment, all the process skills listed in Figure 4 are included in the content of both the DD&E instructional and the DD&E assessment resources.

When we approached the task of designing a device to organize information about student entry-level competence status, it was thus natural that we turned to the matrix of contexts by skills. The matrix provides a convenient and useful means (a) for organizing assessment information, (b) for relating this information to training objectives in order to determine the "discrepancy" profile between current and desired levels of competence, and (c) for relating these discrepancies to the instructional resources provided by the DD&E program.

Before examining the more complex task of employing several different sources of information, let us focus solely on the DD&E Diagnostic Instrument.

## C. THE DD&E DIAGNOSTIC INSTRUMENT

Few students entering the DD&E Program can be expected to have anything approaching a clear idea of the scope and organization of the DD&E Competence domain. Indeed, there really isn't a very good consensus among DD&E practitioners. But, both the student and the instructor need some reasonable basis for developing an instructional plan.

The DD&E Diagnostic instrument makes what seems to be a reasonable premise that the student is able to make some rough judgments about what he or she knows or can do with respect to relatively specific tasks, e.g.:

"Write copy for instructional materials from product specifications."

---

*Job and task analyses indicated that competence in these two areas are especially important for entry-professionals. Although familiarization with these skills can be provided with each context series, it is not feasible to provide sufficient attention to these significant skills by this means alone.

"Determine if theoretical assumptions underlying various statistical techniques have been violated in the analysis of data."

"Identify the crucial characteristics of a target group which may influence the dissemination effort."

The student is asked to rate these items in terms of the following scale:*

Level of Proficiency

1. I have no specific knowledge about this activity nor experience with it.
2. I have read about or seen this activity performed, but have no experience with it and don't really understand it.
3. I have studied this activity or have frequently seen it performed and have a good understanding of it, but I have not yet done it.
4. I have a general understanding of this activity and have had some experience with it, enough so that I can do it if I have either detailed instructions or close supervision.
5. I have enough experience in performing this activity to do it if given enough general supervision or general instructions.
6. I have enough knowledge and experience with this activity so that I can perform this task quite satisfactorily without supervision or job aids.
7. I have had extensive experience with this activity, and can perform it quickly, efficiently and do a top quality job.

The DD&E Diagnostic Instrument consists of 72 items that are rated on this scale. The items are organized in blocks with content reflecting the matrix structure suggested in Figure 4, with three sets of 24 items each for the context areas of (1) development, (2) field test and evaluation, and (3) dissemination and marketing.

---

*In various field tests, we have used 5-, 6-, and 7-point scales, including scales that treat knowledge of and proficiency in performing separately. Because ratings of knowledge and of performance have displayed extremely high correlations (e.g., .9) there is some justification for the confounding that occurs in this scale.

## Figure 5

# DD&E DIAGNOSTIC INSTRUMENT RECORD FORM

(Original Form)

| | DEVELOPMENT | FIELD TEST & EVALUATION | DISSEMINATION & MARKETING | | |
|---|---|---|---|---|---|
| COLLECTING INFORMATION | [ ] 1. Obtain information on problems<br>[ ] 2. Locate strategies for development<br>[ ] 3. Prepare a search strategy<br>[ ] 4. Evaluate information for relevance<br>[ ] TOTAL AVERAGE [ ] | [ ] 25. Prepare a coding scheme<br>[ ] 26. Organize statistical data<br>[ ] 27. Interpret scatter plots<br>[ ] 28. Insure privacy in data collection<br>[ ] TOTAL AVERAGE [ ] | [ ] 49. Collect data on installation costs<br>[ ] 50. Design marketing study data forms<br>[ ] 51. Use information services<br>[ ] 52. Construct an annotated bibliography<br>[ ] TOTAL AVERAGE [ ] | ROW TOTAL | ROW AVERAGE* |
| ANALYZING | [ ] 5. Write a 20 page proposal<br>[ ] 6. Use instructional theories in design<br>[ ] 7. Review field test data for revision<br>[ ] 8. Classify instructional objectives<br>[ ] TOTAL AVERAGE [ ] | [ ] 29. Discuss internal & external validity<br>[ ] 30. Formulate evaluation questions<br>[ ] 31. Analyze test outcomes<br>[ ] 32. Decide if tests fit evaluation plan<br>[ ] TOTAL AVERAGE [ ] | [ ] 53. Indentify character of target group<br>[ ] 54. Evaluate market research techniques<br>[ ] 55. Determine problems in installation<br>[ ] 56. Secure copyright clearances<br>[ ] TOTAL AVERAGE [ ] | ROW TOTAL | ROW AVERAGE* |
| PLANNING | [ ] 9. Plan a budget<br>[ ] 10. Specify format of materials<br>[ ] 11. Sequence learning activities<br>[ ] 12. Estimate needed production materials<br>[ ] TOTAL AVERAGE [ ] | [ ] 33. Determine validity of your test<br>[ ] 34. Design a project monitoring system<br>[ ] 35. Plan control of extraneous variables<br>[ ] 36. Set criteria for field test sites<br>[ ] TOTAL AVERAGE [ ] | [ ] 57. List important dissemination factors<br>[ ] 58. Design a public relations activity<br>[ ] 59. Determine a market for your product<br>[ ] 60. Prepare specifications for A/V material<br>[ ] TOTAL AVERAGE [ ] | ROW TOTAL | ROW AVERAGE* |
| PRODUCING | [ ] 13. Secure waivers, releases, etc.<br>[ ] 14. Confer with specialists when needed<br>[ ] 15. Write instructional exercises<br>[ ] 16. Write copy from specifications<br>[ ] TOTAL AVERAGE [ ] | [ ] 37. Use evaluation data to revise tests<br>[ ] 38. Adjust test procedures when needed<br>[ ] 39. Prepare a test administration manual<br>[ ] 40. Discuss standardized interviews<br>[ ] TOTAL AVERAGE [ ] | [ ] 61. Write scripts for film & slide shows<br>[ ] 62. Write press releases about a product<br>[ ] 63. Prepare a sample product<br>[ ] 64. Write specifications for user manual<br>[ ] TOTAL AVERAGE [ ] | ROW TOTAL | ROW AVERAGE* |
| EVALUATING | [ ] 17. Make recommendations from field data<br>[ ] 18. Conduct a case study of a program<br>[ ] 19. Check product against specifications<br>[ ] 20. Informally try out a product<br>[ ] TOTAL AVERAGE [ ] | [ ] 41. Decide if statistics are suitable<br>[ ] 42. Evaluate test instruments<br>[ ] 43. Find internal consistency of a test<br>[ ] 44. Identify an evaluation's purpose<br>[ ] TOTAL AVERAGE [ ] | [ ] 65. Evaluate the product's distribution<br>[ ] 66. Evaluate effect of a demonstration<br>[ ] 67. Identify parts of a marketing study<br>[ ] 68. Interview users about product use<br>[ ] TOTAL AVERAGE [ ] | ROW TOTAL | ROW AVERAGE* |
| COMMUNICATING | [ ] 21. Write a position paper<br>[ ] 22. Discuss a product with user groups<br>[ ] 23. Interact in a staff meeting<br>[ ] 24. Communicate product specifications<br>[ ] TOTAL AVERAGE [ ] | [ ] 45. Prepare an article for publication<br>[ ] 46. Give a short speech or oral report<br>[ ] 47. Prepare simple evaluation reports<br>[ ] 48. Prepare graphs to display data<br>[ ] TOTAL AVERAGE [ ] | [ ] 69. Establish cooperation with users<br>[ ] 70. Deliver an oral presentation<br>[ ] 71. Demonstrate the use of a product<br>[ ] 72. Translate data into verbal form<br>[ ] TOTAL AVERAGE [ ] | ROW TOTAL | ROW AVERAGE* |
| | [ ] COLUMN TOTAL COLUMN AVERAGE [ ]* | [ ] COLUMN TOTAL COLUMN AVERAGE [ ]* | [ ] COLUMN TOTAL COLUMN AVERAGE [ ]* | | |

*Column Total ÷ 24 = Column Average

*Row Total ÷ 12 = Row Average

The instrument record form, reproduced in reduced size, appears in Figure 5. Content of the items is abbreviated on the record form but does provide the student and instructor with a convenient and specific reference to the content of each item. The form provides for the computation of averages for cells, rows and columns. The record form thus provides a compact and immediately interpretable summary of the student's self-assessment for a systematically constructed sample of DD&E tasks.

The record form can provide a highly useful basis for the student and instructor to discuss:

- the student's perception of current level of competence
- which appear to be the areas of greatest and least competence and what this may mean in terms of the student's training, experience, and aspirations
- explanations for any unusual patterns (e.g., items or sets of items that have been rated particularly low or high relative to (1) the student's overall profile of ratings, (2) patterns seen in other students with similar education or experience, or (3) what would be expected given specific information about the student's training or experience
- what the student perceives as desired levels of competence in each context or skill area given the student's instruction program objectives
- what the instructor would advise as to desired levels of competence
- which competence areas deserve special attention in planning the student's program of instruction

Now, for some comments. First, please keep in mind that the Diagnostic Instrument Record form is only a point of departure for establishing a "Bayesian" competence record. We have used the label "Bayesian" in quotations to signify that while we are not asking the student or the instructor to employ actual Bayesian computations (See, for example, Novick and Jackson, 1974.) we are asking both to employ a weak form of Bayesian inference.

In making any decision about the student's level of competence, other than in the case where it can be referred to a highly specific, well-defined, readily and reliably measurable, unequivocably "valid," and context invariant performance (which is highly unlikely), we have to make inferences. Although it is a gross oversimplification, the thrust of the Bayesian approach is to incorporate "prior" (previously available) and "collateral" (concurrently available) information with specific "current" information in making an inference (e.g., regarding individual proficiency level in some DD&E skill or set of skills) rather than basing the inference on the "current" data only.

One major purpose of the DD&E Assessment System is to provide both the student and the instructor with relatively simple means for collecting, organizing, and employing such prior and collateral information at each significant decision point in the student's instruction program.

Consider for the moment the data that might be relevant to the record in Figure 5. Assume that as an instructor you have worked with several dozen students discussing the meaning and interpretation of such records, and then assume that you are discussing the record with a particular student. You would have several important kinds of prior information including: (1) what the self ratings of other students looked like, (2) other information such as courses taken, grades, work experience, course performance, etc., on these other students, and (3) similar information on courses, grades, etc. for this student. This prior information can and should be employed in asking whether the student's ratings are consistent with this prior information. Also, assuming some relationship among the skills, it is also possible to use the collateral information contained in the other items or averages for other groups of items. Thus, if the student self-rated relatively low on most items, and this was consistent with the other background information on this student and relative to other students, but the student self rated much higher in one area, one might suspect that this rating was due to the unusual interpretation the student made regarding the items, to carelessness in marking, or perhaps to relatively higher competence in a particular item or set of items. In any case, the situation would call for investigation which might include any of the following:

- asking the student to discuss the reasons for rating the items at the recorded levels
- probing the student for relevant information on past experience or training that would confirm or disconfirm the ratings

- asking the student to rate additional items (which may be available in the item pool or which could be prepared by the instructor) reflecting related tasks in the same general competence area
- asking the student to complete several knowledge test items from the knowledge test item pool (available for most of the areas)

If the results tended to confirm the anomalous rating, one could have reasonable confidence that the student's self-rating was correct.

Obviously, a similar procedure would be indicated in the event that: (a) the overall level of reported proficiency was unexpectedly high or low, (b) specific competence areas (sets of related items) were unexpectedly high or low.

We are not suggesting that these procedures be drawn out into a protracted sequence of sequential testing which could easily exhaust the patience of both the student and the instructor. But we are trying to make at least three points:

- The DD&E assessment system provides methods and resources to make follow-up assessment when this may be useful.
- The task of interpreting the Diagnostic Instrument data should be approached with such a "Bayesian" perspective about the use of all reasonably relevant data in making any initial inferences about student competence.
- This general method should be employed in the subsequent use of new assessment information that will enter students' files as they progress in their DD&E training programs.

## D. VALIDATION TEST OF THE DIAGNOSTIC INSTRUMENT

### 1. The Field Test

Tables 1, 2 and 3 summarize validation data collected on a total of 78 subjects representing a wide range of experience and training. Table 1 describes the composition of the test population. Table 2 reports item means and significance tests for each of the 72 items. Table 3 reports scale means, standard deviations, item means (scale means divided by number of items in the scale; these item means permit direct comparison among scales and to the proficiency level rating descriptions), correlations among scales, and factor loadings for:

TABLE 1

## DD&E DIAGNOSTIC INSTRUMENT VALIDATION SUBJECTS

| Number | Description of Validation Subjects* |
|---|---|
| 11 | Students recruited from a local university whose backgrounds were similar to that of students we expect to enter the DD&E program. They were seniors or graduate students majoring in education or majoring in the behavioral sciences with an interest in education. |
| 51 | Graduate Students (Masters and Ph.D. candidates) enrolled in educational research and development courses. These subjects were recruited by their professors at Temple University, San Francisco State University, Michigan State University, and UCLA. |
| 12 | Developers - Masters level personnel with at least one year of experience in educational development, or persons with Ph.D.'s in education who have some association with educational development or who are presently developing educational products. Seven of this group were enrolled in inservice training in DD&E. |
| 4 | Senior Developers - Highly experienced in developing educational products and programs. These personnel are presently working in the field. They have published in the field and/or produced products that have been distributed and are being used. |
| 78 | TOTAL |

---

*Note analysis of variance indicated that there were no substantial differences between students and graduate students so these groups were combined and are identified as students in Table 2. Although senior developers tended to have higher competence ratings than developers, the number of senior developers was too small to demonstrate statistical significance. These two groups were combined and are identified as developers in Table 2.

(a) the three *context* scales (i.e., column totals of figure 5), (b) the *six process* skill scales (i.e., row totals of Figure 5), and (c) four empirically derived cluster scales (see Appendix A . Please note that this group of 78 subjects is predominantly a student population, which is appropriate since the Diagnostic Instrument is intended for professional education settings. The number of experienced developers is small (N=16) but sufficient to demonstrate the ability of the DD&E instrument to discriminate between students and developers.

Apendix A contains a complete report of the data analysis and a copy of the instrument. The following pages summarize the findings.

## 2. *Content of the Instrument*

Analysis of the data indicated that the DD&E Diagnostic Instrument items were highly intercorrelated but that at least three or possibly four subscales can be differentiated. Cluster analysis resulted in four well-defined clusters, three associated with the DD&E contexts and a fourth derived primarily from items appearing in the D&D contexts relating to publication, production, and public relations. However, correlation and factor analysis indicate that this last cluster is highly correlated with and exhibits a factor pattern similar to the dissemination cluster. The factor analysis results suggest that only three factor scales are needed: (1) development, (2) field test and evaluation, and (3) production, dissemination, and marketing.

Neither the cluster analysis nor the factor analysis resulted in a group of competence statements associated primarily with any *one* of the process skills; however, the skill scale factor loadings on the newly defined DD&E factors *exhibit different patterns of skills* loadings which appear to be meaningful. The *evaluation factor* is most prominently associated with: collecting information, analyzing, and evaluating; the *production, dissemination and marketing factor* with: planning and designing, producing, and communicating; the *development factor* with: analyzing, planning and designing, producing, and evaluating.

The analysis indicated that the items which had been placed *a priori* in the development *context*, were in fact a mixture of all three DD&E competencies, consequently, the number of items associated purely with development is closer to nine in number rather than the intended 24.

## TABLE 2

### DIAGNOSTIC INSTRUMENT ITEM MEANS AND SIGNIFICANCE TESTS

| ABBREVIATED ITEM CONTENT | MEANS STUDENTS | MEANS DEVELOPERS | F-RATIO | p-LEVEL |
|---|---|---|---|---|
| DEVELOPMENT | (N=62) | (N=16) | | |
| 1. Obtain information on problems | 3.97 | 5.62 | 9.49 | .003 |
| 2. Locate strategies for development | 3.73 | 5.44 | 11.88 | .001 |
| 3. Prepare a search strategy | 4.42 | 5.88 | 8.50 | .005 |
| 4. Evaluate information for relevance | 5.35 | 6.44 | 7.55 | .008 |
| 5. Write a 20 page proposal | 4.06 | 5.00 | 5.04 | .026 |
| 6. Use instructional theories in design | 3.87 | 5.75 | 16.73 | .000 * |
| 7. Review field test data for revision | 3.44 | 4.94 | 8.86 | .004 |
| 8. Classify instructional objectives | 4.03 | 5.50 | 6.59 | .012 |
| 9. Plan a budget | 3.50 | 4.56 | 4.02 | .046 |
| 10. Specify format of materials | 3.68 | 5.31 | 9.40 | .003 |
| 11. Sequence learning activities | 4.37 | 5.19 | 2.70 | .101 NS |
| 12. Estimate needed production materials | 3.81 | 5.69 | 11.59 | .001 |
| 13. Secure waivers, releases, etc. | 2.58 | 3.94 | 6.69 | .011 |
| 14. Confer with specialists when needed | 3.92 | 5.19 | 4.90 | .028 |
| 15. Write instructional exercises | 4.53 | 5.56 | 3.96 | .048 |
| 16. Write copy from specifications | 3.76 | 4.75 | 3.04 | .081 NS |
| 17. Make recommendations from field data | 3.85 | 5.50 | 13.37 | .001 |
| 18. Conduct a case study of a program | 3.95 | 5.00 | 4.45 | .036 |
| 19. Check product against specifications | 3.76 | 5.38 | 10.31 | .002 |
| 20. Informally try out a product | 4.55 | 6.19 | 12.03 | .001 |
| 21. Write a position paper | 3.92 | 5.00 | 4.27 | .040 |
| 22. Discuss a product with user groups | 4.03 | 5.62 | 9.90 | .003 |
| 23. Interact in a staff meeting | 5.21 | 6.38 | 8.41 | .005 |
| 24. Communicate product specifications | 3.58 | 5.62 | 15.14 | .000 |
| TOTAL - DEVELOPMENT | 3.99 | 5.39 | 20.12 | .0001 |

* - All Values tabled .000 have p-Levels of .0005 or less

NS - Not significant at 5% level

TABLE 2 (cont'd.)

| ABBREVIATED ITEM CONTENT<br>FIELD TEST & EVALUATION | MEANS STUDENTS (N=62) | MEANS DEVELOPERS (N=16) | F-RATIO | p-LEVEL |
|---|---|---|---|---|
| 25. Prepare a coding scheme | 2.94 | 4.19 | 4.45 | .036 |
| 26. Organize statistical data | 4.31 | 5.62 | 7.09 | .009 |
| 27. Interpret scatter plots | 3.65 | 4.81 | 3.98 | .047 |
| 28. Insure privacy in data collection | 3.76 | 5.25 | 9.10 | .004 |
| 29. Discuss internal & external validity | 4.10 | 5.06 | 4.17 | .042 |
| 30. Formulate evaluation questions | 3.98 | 5.19 | 7.63 | .007 |
| 31. Analyze test outcomes | 3.85 | 5.50 | 13.94 | .001 |
| 32. Decide if tests fit evaluation plan | 3.42 | 5.31 | 21.05 | .000 |
| 33. Determine validity of your test | 3.13 | 4.56 | 10.98 | .002 |
| 34. Design a project monitoring system | 2.56 | 4.88 | 21.96 | .000 |
| 35. Plan control of extraneous variables | 3.92 | 4.69 | 2.62 | .106 NS |
| 36. Set criteria for field test sites | 3.37 | 4.94 | 9.99 | .003 |
| 37. Use evaluation data to revise tests | 3.95 | 5.69 | 16.16 | .000 |
| 38. Adjust test procedures when needed | 3.48 | 5.00 | 9.20 | .004 |
| 39. Prepare a test administration manual | 3.15 | 5.06 | 15.82 | .000 |
| 40. Discuss standardized interviews | 4.44 | 5.62 | 6.42 | .013 |
| 41. Decide if statistics are suitable | 3.50 | 4.44 | 3.91 | .049 |
| 42. Evaluate test instruments | 3.60 | 5.56 | 16.04 | .000 |
| 43. Find internal consistency of a test | 3.56 | 4.69 | 5.63 | .019 |
| 44. Identify an evaluation's purpose | 3.76 | 5.62 | 14.96 | .000 |
| 45. Prepare an article for publication | 4.15 | 5.06 | 3.35 | .068 NS |
| 46. Give a short speech or oral report | 6.02 | 6.38 | 1.24 | .269 NS |
| 47. Prepare simple evaluation reports | 4.87 | 6.25 | 11.21 | .002 |
| 48. Prepare graphs to display data | 5.08 | 5.88 | 2.77 | .096 NS |
| TOTAL - EVALUATION | 3.86 | 5.22 | 19.59 | .0001 |

* - All Values tabled .000 have p-Levels of .0005 or less

NS - Not significant at 5% level

TABLE 2 (cont'd.)

| ABBREVIATED ITEM CONTENT | MEANS STUDENTS | MEANS DEVELOPERS | F-RATIO | p-LEVEL |
|---|---|---|---|---|
| DISSEMINATION & MARKETING | (N=62) | (N=16) | | |
| 49. Collect data on installation costs | 2.95 | 4.25 | 5.74 | .018 |
| 50. Design marketing study data forms | 2.79 | 3.69 | 3.75 | .054 NS |
| 51. Use information services | 3.69 | 5.81 | 16.22 | .000 |
| 52. Construct an annotated bibliography | 4.55 | 5.12 | 1.05 | .310 NS |
| 53. Identify character of target group | 3.76 | 4.62 | 2.76 | .097 NS |
| 54. Evaluate market research techniques | 2.34 | 3.69 | 8.17 | .006 |
| 55. Determine problems in installation | 3.08 | 5.00 | 15.18 | .000 |
| 56. Secure copyright clearances | 2.65 | 3.75 | 4.05 | .045 |
| 57. List important dissemination factors | 3.53 | 4.25 | 1.71 | .191 NS |
| 58. Design a public relations activity | 2.71 | 3.50 | 2.18 | .140 NS |
| 59. Determine a market for your product | 3.34 | 4.31 | 3.26 | .072 NS |
| 60. Prepare specifications for A/V material | 3.02 | 3.75 | 2.12 | .146 NS |
| 61. Write scripts for film & slide shows | 3.19 | 3.94 | 1.84 | .176 NS |
| 62. Write press releases about a product | 3.16 | 3.69 | 0.95 | .666 NS |
| 63. Prepare a sample product | 3.56 | 5.06 | 9.34 | .003 |
| 64. Write specifications for user manual | 3.74 | 4.69 | 3.28 | .070 NS |
| 65. Evaluate the product's distribution | 2.77 | 3.75 | 3.77 | .053 NS |
| 66. Evaluate effect of a demonstration | 3.77 | 4.69 | 3.80 | .052 NS |
| 67. Identify parts of a marketing study | 3.37 | 4.06 | 1.55 | .214 NS |
| 68. Interview users about product use | 3.89 | 5.12 | 5.97 | .016 |
| 69. Establish cooperation with users | 4.15 | 5.50 | 7.34 | .008 |
| 70. Deliver an oral presentation | 4.11 | 5.38 | 5.58 | .020 |
| 71. Demonstrate the use of a product | 4.53 | 5.19 | 1.64 | .202 NS |
| 72. Translate data into verbal form | 3.81 | 4.38 | 1.06 | .308 NS |
| TOTAL - DISSEMINATION | 3.44 | 4.47 | 8.41 | .00051 |

* - All Values tabled .000 have p-Levels of .0005 or less

NS - Not significant at 5% level

## TABLE 3

## DD&E DIAGNOSTIC INSTRUMENT - MEANS, STANDARD DEVIATIONS, ITEM MEANS, AND FACTOR ANALYSIS OF CONTEXT, PROCESS SKILL, AND CLUSTER SCALES

Decimals omitted for correlations and factor leadings

(N=78)

| | Scale Mean | Scale S.D. | Item Mean | Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | Variable | Factor I | Factor II | Factor III | Factor IV | % C. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Development Context | 102.8 | 29.6 | 4.28 | 1 | - | 76 | 79 | 84 | 90 | 90 | 89 | 86 | 86 | 93 | 80 | 84 | 79 | 94 | 45 | 1 | 59 | 50 | 56 | 21 | 96 |
| Field Test and Evaluation Context | 99.2 | 29.2 | 4.14 | 2 | 76 | - | 58 | 85 | 84 | 77 | 71 | 87 | 76 | 65 | 99 | 64 | 61 | 86 | 36 | 2 | 20 | 91 | 32 | 12 | 98 |
| Dissemination and Marketing Context | 87.5 | 31.6 | 3.65 | 3 | 79 | 58 | - | 72 | 80 | 90 | 85 | 79 | 87 | 70 | 56 | 98 | 88 | 89 | 31 | 3 | 28 | 25 | 90 | 11 | 97 |
| Collecting Information Skills | 49.4 | 14.4 | 4.12 | 4 | 84 | 85 | 72 | - | 84 | 81 | 71 | 81 | 81 | 71 | 85 | 77 | 73 | 89 | 40 | 4 | 20 | 72 | 52 | 18 | 86 |
| Analyzing Skills | 46.0 | 14.4 | 3.83 | 5 | 90 | 84 | 80 | 84 | - | 90 | 83 | 89 | 80 | 84 | 85 | 84 | 76 | 94 | 43 | 5 | 45 | 62 | 54 | 20 | 91 |
| Planning Skills | 43.9 | 14.8 | 3.66 | 6 | 90 | 77 | 90 | 81 | 90 | - | 88 | 87 | 87 | 81 | 77 | 93 | 84 | 96 | 39 | 6 | 40 | 48 | 72 | 15 | 93 |
| Producing Skills | 46.5 | 15.4 | 3.87 | 7 | 89 | 71 | 85 | 71 | 83 | 88 | - | 81 | 84 | 83 | 72 | 86 | 84 | 91 | 36 | 7 | 50 | 39 | 68 | 13 | 88 |
| Evaluating Skills | 47.6 | 14.4 | 3.96 | 8 | 86 | 87 | 79 | 81 | 89 | 87 | 81 | - | 83 | 78 | 89 | 83 | 69 | 93 | 39 | 8 | 38 | 69 | 51 | 14 | 89 |
| Communicating Skills | 56.2 | 14.3 | 4.68 | 9 | 86 | 76 | 87 | 81 | 80 | 87 | 84 | 83 | - | 73 | 74 | 89 | 85 | 92 | 34 | 9 | 23 | 50 | 76 | 11 | 89 |
| Development Cluster | 38.7 | 12.6 | 4.30 | 10 | 93 | 65 | 70 | 71 | 84 | 81 | 83 | 78 | 73 | - | 69 | 76 | 62 | 85 | 41 | 10 | 79 | 37 | 41 | 18 | 98 |
| Field Test and Evaluation Cluster | 102.6 | 30.9 | 4.10 | 11 | 80 | 99 | 56 | 85 | 85 | 77 | 72 | 89 | 74 | 69 | - | 63 | 58 | 86 | 38 | 11 | 28 | 90 | 27 | 15 | 99 |
| Dissemination Cluster | 93.5 | 31.5 | 3.89 | 12 | 84 | 64 | 98 | 77 | 84 | 93 | 86 | 83 | 89 | 76 | 63 | - | 83 | 92 | 37 | 12 | 34 | 32 | 84 | 15 | 94 |
| Publication, Production and P.R. Cluster | 43.4 | 16.3 | 3.62 | 13 | 79 | 61 | 88 | 73 | 76 | 84 | 84 | 69 | 85 | 62 | 58 | 83 | - | 85 | 30 | 13 | 15 | 31 | 87 | 10 | 89 |
| TOTAL SCORE | 289.5 | 81.2 | 4.02 | 14 | 94 | 86 | 89 | 89 | 94 | 96 | 91 | 93 | 92 | 85 | 86 | 92 | 85 | - | 41 | 14 | 39 | 61 | 67 | 16 | 100 |
| Group 1=St, 2=G. St, 3=Dev, 4=Sr. Dev. | 2.12 | .69 | 2.12 | 15 | 45 | 36 | 31 | 40 | 43 | 39 | 36 | 39 | 34 | 41 | 38 | 37 | 30 | 41 | - | 15 | 13 | 18 | 14 | 97 | 100 |
| Percent of trace | | | | | | | | | | | | | | | | | | | | | 16 | 31 | 39 | 08 | - |

### 3. Discrimination Validity

Given the predominantly student population of the field test sample (62 of 78 subjects), the average competence level and the distribution of item means reported in Table 3 seems acceptable, since the average proficiency level for the entire instrument is at the midpoint (4.0) on the seven-point scale, and the great majority of the item proficiency levels are between 3.0 and 5.0 (61 of the 72 items). Analysis of variance between the student and the developer groups indicates that the majority of the items and all of the scales discriminate between these groups. In terms of the proficiency scale, the average difference is 1.27 proficiency levels. The means for the majority of the competence statements for the student group were below 4.0 which corresponds to the statement:

> "I have a general understanding of this activity and have had some experience with it, enough so I can do it if I have either detailed instructions or close supervision."

On the other hand, the majority of the competence statement means for developers were above 5.0, which corresponds to the statement:

> "I have enough experience in performing this activity to do it if given enough general supervision or general instructions."

This clear, meaningful, and highly significant difference ($p < .0001$) provides further evidence that the instrument is able to discriminate between groups with different amounts of relevant experience. (Previous pilot studies have shown that self ratings made by educational developers correlated with supervisors' ratings of the same sets of competencies, .73 for one scale and .88 for another, thus indicating that experienced developers and their supervisors exhibit substantial agreement regarding the average levels of proficiency.)

In terms of levels of self-rated proficiency, both the students and the developers reported relatively higher levels of proficiency in development, followed closely by field test and evaluation, and relatively lower levels of proficiency in publication, production, and public relations competencies. Both groups rated communication skills highest and planning and designing skills lowest. The difference between the two groups was at least one proficiency level for all subscales (contexts, skills, cluster analysis scales), with the greatest differences appearing on the development context scale, the field test and evaluation cluster scale and on the analyzing, the collecting information, and the evaluating skills scales. The smallest differences between the two groups were on the dissemination and marketing context scale, the (related) publication, production, and public relations cluster scale, and the communication skills scale. In the case of communication, the relatively small diference in proficiency level seems to be attributable primarily to a (relatively) high level of proficiency in the student group with respect to oral and written communication. In the case of the dissemination and marketing context scale and the publication, production, and marketing cluster scale the smaller difference between groups is primarily attributable to the relatively lower levels of proficiency reported by the developers.

### 4. Revisions

Given the results of the data analysis, it was decided to retain the content structure of the instrument (matrix of three contexts by six process skills), but to revise a number of items so that the three DD&E factors would be more clearly and evenly represented and more closely associated with the three context columns of the response form. In addition, an eighth level was added to the proficiency rating scale to separate the expert from the highly experienced. Appendix A includes copies of both the original and the revised instrument.

## E. INTERPRETATION AND USE OF THE DIAGNOSTIC INSTRUMENT IN STUDENT PROGRAM PLANNING

Several field tests of various versions of the DD&E Diagnostic Instrument have indicated that the instrument is capable of discriminating between inexperienced students and experienced DD&E personnel. Moreover, in those cases where we have been able to obtain supervisor's ratings for employed DD&E professionals, self ratings and supervisor ratings have displayed substantial correlations (e.g. , .73 and .88). The individual items and the subscales display substantial intercorrelations, but factor analysis and cluster analysis indicate that at least three different factors (or clusters) can be identified, corresponding to the three DD&E contexts. Neither the cluster analysis nor the factor analysis provide evidence for separate skills factors or clusters; however, the factor analysis indicates that the three DD&E factors (Factors I, II and III respectively in Table 3) display loadings with different patterns of skills for DD&E.

The data in Table 2 indicate that the average student tends to rate most items "3" (good understanding but have not yet done it) or "4" (can do it with detailed instructions or close supervision). However these data include a substantial number of students who had completed or were enrolled in educational R&D courses. Entering students may be expected to display somewhat lower average ratings. Obviously, our best advice is for the instructor to compile rating data based on students actually entering his or her own program and to use these data in evaluating the DD&E Diagnostic Instrument Ratings of individual students.

The instructor should keep in mind, and explain to the student, that the 72 items contained in the DD&E Diagnostic Instrument are but samples of a much larger potential population of DD&E competence items. Hence, caution should be exercised in placing too much imphasis on an individual item, or even on the average of the set of four items appearing in each "cell" of Figure 5 in terms of making inferences about the student's competence; however, the totals (or averages) for columns (based on 24 items) and rows (based on 12 items) are probably reliable indicators of the student's general level of competence in the specified area.

Regarding the validity of the self rating data, Table 2 clearly indicates that we can make useful and meaningful discriminations between groups of persons with different levels of experience. Needless to say, individuals will vary in terms of their tendency to under- or overrate their competence, and quite possibly in their interpretation of the scale descriptions and the meaning of the competence items themselves. Hence, it is essential that the instructor consider both prior and concurrent information about the student and about the student's peers, in assessing and evaluating the Diagnostic Instrument ratings as information about attained competence levels.

As its label implies, the instrument is intended for diagnostic purposes. It provides a useful device for reviewing and discussing with the student not only the student's perception of current levels of competence, but the student's ideas about the levels of competence which should be set as objectives at the completion of the student's program of instruction.

In terms of objectives for student program planning, the item means and totals reported in Table 2 for developers provide some guidance. With some exceptions, it would seem that a "4" rating (I have a general understanding of this activity and have had some experience with it, enough so I can do it if I have either detailed instructions or close supervision) or a "5" rating (I have enough experieince in performing this activity to do it if given enough general supervision or general instructions) would be a reasonable objective (in an "entry level" i.e. M.A. degree program) for most, but not all, competence items. Certainly the data in Table 2 suggest that the entry level DD&E student should aim for rating averages of at least "5.0" in Development and in Field Test and Evaluation. But note that the developers, as a group, rated themselves below "5" on 16 of 24 items in Dissemination and Marketing context. Despite this empirical self-report evidence, indicating that experienced developers average only "4.5" in this context area, an average of "5.0" would be a desirable dissemination and marketing competence objective for any student aiming for all round general competence in DD&E.

Needless to say, students should be encouraged to define and justify their own particular sets of competence objectives, not only in terms of the sample of items provided by the DD&E Diagnostic Instrument, but perhaps by writing their own statements of competencies they wish to obtain. In terms of levels of proficiency, there may be specific terminal competence items at levels "3" (studied but have not done it), "4" (can do with detailed instructions or close supervision), "5" (can do with general supervision or instructions), and "6" can perform satisfactorily without supervision or job aids). Perhaps in specific instances, there may even be items which could be justified at levels "1" (no knowledge) or "2" (don't really understand).

The important thing in a competence based, individualized program of instruction, is that the student and the instructor develop some tentative, but explicit, conception about both the individual student's entry level and the desirable exit level for that student. The magnitudes of the discrepancies between these two levels for different sets of competencies can then be used to help define the areas in which knowledge or skill attainment is needed.

If not taken too literally, the form displayed in Figure 5 can be used as a device that can be employed to record the student's preliminary ideas about terminal objectives and also to identify the magnitude of the discrepancy between current and desired competence levels. This can be accomplished by entering an exit level objective expressed as a rating scale number beside each average and then computing the differences between the recorded average and the student's objective.

Figure 6 gives an (real) example for a particular student. In this case the objective (Obj.) for each cell (context by skill combination) is indicated immediately to the right of the Average and is followed by the discrepancy (D) between the self-rating average and the objective. Most values have been rounded to one decimal. Major discrepancies are circled. This particular student is relatively low on self-rating for analyzing skills (2.8), planning skills (2.5), and evaluating skills (2.7) with several of the competence items rated "1." In terms of the three DD&E contexts, field test and evaluation is highest (3.8), followed by development (3.4), and then dissemination and marketing (2.5). When compared to the student item means reported in Table 2, we see that the student's average for field test and evaluation is quite close to that reported in Table 2, the student's average for development is about .6 point lower, but the student's average for dissemination and marketing is nearly a full point lower, which is probably a significant difference.

## FIGURE 6 - EXAMPLE OF USE OF DD&E DIAGNOSTIC INSTRUMENT RESPONSE FORM AS A STUDENT PROGRAM PLANNING DEVICE

| | DEVELOPMENT | FIELD TEST & EVALUATION | DISSEMINATION & MARKETING | | |
|---|---|---|---|---|---|
| COLLECTING INFORMATION | [4] 1. Obtain information on problems<br>[3] 2. Locate strategies for development<br>[3] 3. Prepare a search strategy<br>[6] 4. Evaluate information for relevance<br>[16] TOTAL AVERAGE [4.0] Obj. = 5.0 D=1.0 | [1] 25. Prepare a coding scheme<br>[6] 26. Organize statistical data<br>[5] 27. Interpret scatter plots<br>[6] 28. Insure privacy in data collection<br>[18] TOTAL AVERAGE [4.5] Obj=5.5. D=1.0 | [2] 49. Collect data on installation costs<br>[1] 50. Design marketing study data forms<br>[1] 51. Use information services<br>[6] 52. Construct an annotated bibliography 4th<br>[10] TOTAL AVERAGE [2.5] Obj=5.0 D=2.5 | ROW TOTAL<br>44 | Obj=5.2<br>D=1.5<br>ROW AVERAGE<br>3.7 |
| ANALYZING | [6] 5. Write a 20 page proposal<br>[5] 6. Use instructional theories in design<br>[2] 7. Review field test data for revision<br>[3] 8. Classify instructional objectives<br>[16] TOTAL AVERAGE [4.0] Obj. = 5.0 D=1.0 | [3] 29. Discuss internal & external validity<br>[2] 30. Formulate evaluation questions<br>[5] 31. Analyze test outcomes<br>[2] 32. Decide if tests fit evaluation plan<br>[12] TOTAL AVERAGE [3.0] Obj=5.0 D=2.0 | [1] 53. Indentify character of target group<br>[1] 54. Evaluate market research techniques<br>[2] 55. Determine problems in installation 1st<br>[2] 56. Secure copyright clearances<br>[6] TOTAL AVERAGE [1.5] Obj = 4.0 = 3.5 | ROW TOTAL<br>34 | Obj=4.7<br>D=1.9<br>ROW AVERAGE<br>2.8 |
| PLANNING | [2] 9. Plan a budget<br>[1] 10. Specify format of materials<br>[5] 11. Sequence learning activities 4th<br>[3] 12. Estimate needed production materials<br>[11] TOTAL AVERAGE [2.75] Obj. = 5.25 D=2.5 | [4] 33. Determine validity of your test<br>[1] 34. Design a project monitoring system<br>[5] 35. Plan control of extraneous variables 3rd<br>[1] 36. Set criteria for field test sites<br>[11] TOTAL AVERAGE [2.75] Obj = 5.5 D=2.75 | [1] 57. List important dissemination factors<br>[2] 58. Design a public relations activity<br>[2] 59. Determine a market for your product 2nd<br>[3] 60. Prepare specifications for A/V material<br>[8] TOTAL AVERAGE [2.0] Obj = 5.0 D=3.0 | ROW TOTAL<br>30 | Obj=5.2<br>D=2.7<br>ROW AVERAGE<br>2.5 |
| PRODUCING | [2] 13. Secure waivers, releases, etc.<br>[6] 14. Confer with specialists when needed<br>[6] 15. Write instructional exercises<br>[2] 16. Write copy from specifications<br>[14] TOTAL AVERAGE [3.5] Obj. = 5.0 D=1.5 | [4] 37. Use evaluation data to revise tests<br>[4] 38. Adjust test procedures when needed<br>[3] 39. Prepare a test administration manual<br>[6] 40. Discuss standardized interviews<br>[17] TOTAL AVERAGE [4.25] Obj = 4.25 D=0 | [5] 61. Write scripts for film & slide shows<br>[3] 62. Write press releases about a product<br>[2] 63. Prepare a sample product<br>[2] 64. Write specifications for user manual<br>[12] TOTAL AVERAGE [3.0] Obj = 5.0 D=2.0 | ROW TOTAL<br>43 | Obj=4.8<br>D= 1.2<br>ROW AVERAGE<br>3.6 |
| EVALUATING | [4] 17. Make recommendations from field data<br>[3] 18. Conduct a case study of a program<br>[2] 19. Check product against specifications<br>[4] 20. Informally try out a product<br>[13] TOTAL AVERAGE [3.25] Obj = 4.75 D=1.5 | [4] 41. Decide if statistics are suitable<br>[1] 42. Evaluate test instruments<br>[4] 43. Find internal consistency of a test<br>[2] 44. Identify an evaluation's purpose<br>[11] TOTAL AVERAGE [2.75] Obj = 4.0 D=1.25 | [1] 65. Evaluate the product's distribution<br>[1] 66. Evaluate effect of a demonstration<br>[3] 67. Identify parts of a marketing study<br>[3] 68. Interview users about product use<br>[8] TOTAL AVERAGE [2.0] Obj = 4.0 D=2.0 | ROW TOTAL<br>32 | Obj=4.2<br>D=1.5<br>ROW AVERAGE<br>2.7 |
| COMMUNICATING | [5] 21. Write a position paper<br>[2] 22. Discuss a product with user groups<br>[1] 23. Interact in a staff meeting<br>[4] 24. Communicate product specifications<br>[12] TOTAL AVERAGE [3.0] Obj = 5.0 D=2.0 | [6] 45. Prepare an article for publication<br>[6] 46. Give a short speech or oral report<br>[5] 47. Prepare simple evaluation reports<br>[6] 48. Prepare graphs to display data<br>[23] TOTAL AVERAGE [5.75] Obj = 5.75 D=0 | [2] 69. Establish cooperation with users<br>[3] 70. Deliver an oral presentation<br>[6] 71. Demonstrate the use of a product<br>[5] 72. Translate data into verbal form<br>[16] TOTAL AVERAGE [4.0] Obj = 6.0 D=2.0 | ROW TOTAL<br>51 | Obj=5.6<br>1.4<br>ROW AVERAGE<br>4.25 |
| | [82] COLUMN TOTAL COLUMN AVERAGE [3.4]* | [92] COLUMN TOTAL COLUMN AVERAGE [3.8]* | [60] COLUMN TOTAL COLUMN AVERAGE [2.5]* | Row Total+ ÷ 12 = Row Average | |

*Column Total ÷ 24 = Column Average Obj = 5.0 D=1.6

Obj = 5.0 D=1.2

Obj = 4.8 D=2.3

This student displays considerable variance in the rating of individual items with 13 items rated "1," and 11 items rated "6."* The list of high and low items is interesting. Items rated "5" or "6" include (reading across rows):

(Collecting Information)

4. Evaluate information for relevance
26. Organize statistical data
27. Interpret scatter plots
28. Insure privacy in data collection
52. Construct an annotated bibliography

(Analyzing)

5. Write a 20-page proposal
6. Use instructional theories in design
31. Analyze test outcomes

(Planning)

11. Sequence learning activities
35. Plan control of extraneous variables

(Producing)

14. Confer with specialists when needed
15. Write instructional exercises
40. Discuss standardized interviews
61. Write scripts for film and slide shows

(Evaluating)

[none]

(Communicating)

21. Write a position paper
45. Prepare an article for publication
46. Give a short speech or oral report
47. Prepare simple evaluation reports
48. Prepare graphs to display data
71. Demonstrate the use of a product
72. Translate data into verbal form

Apparently this particular student possesses significant skills in writing and communication, in some statistical/data processing areas, and in several competencies dealing with analysis, planning, and production of instructional materials.

---

*The rating variance for this student is considerably greater than that normally encountered, but not entirely unusual. We have deliberately selected a student record with substantial rating variance in order to illustrate the interpretation of the Diagnostic Instrument data.

Conversely there are a number of competencies in which the student claims no specific knowledge ("1") or understanding ("2"), including:

(Collecting Information)

25. Prepare a coding scheme
49. Collect data on installation costs
50. Design marketing study data forms
51. Use information services

(Analyzing)

7. Review field test data for revision [of a product]
30. Formulate evaluation questions
32. Decide if tests fit evaluation plan
53. Identify character of target group
54. Evaluate market research techniques
55. Determine problems in installation
56. Secure copyright clearances

(Planning)

9. Plan a budget
10. Specify format of materials
34. Design a project monitoring system
36. Set criteria for field test sites
57. List important dissemination factors
58. Design a public relations activity
59. Determine a market for your product

(Producing)

13. Secure waivers, released, etc.
16. Write copy from specifications
63. Prepare a sample product
64. Write specifications for users' manual

(Evaluating)

19. Check product against specifications
42. Evaluate test instruments
44. Identify an evaluation's purpose
65. Evaluate the product's distribution
66. Evaluate effect of a demonstration

(Communicating)

22. Discuss a product with user groups
23. Interact in a staff meeting
69. Establish cooperation with users

This appears to be a substantial number of competencies, spread among all three DD&E contexts and all six skill areas, where the student claims to have neither knowledge nor skills.

Now, let us make the very important assumption that the instructor has discussed these ratings with the student and has done enough probing of past training and experience to be willing to accept (at least tentatively) the ratings as reported in Figure 6. Given the student's somewhat lower than average ratings in several competence areas, the student was encouraged to set objectives that would tend to bring most of the competence rating averages more in line and closer to an overall "5.0" level, which would indicate ability to perform most tasks if given general supervision or general instructions.*

Objectives were actually set by considering the four "sample" items appearing in each cell (context and skills). To emphasize the point that these four items are samples, the instructor should urge the student to name another example of a competence that would fit in the same cell. After the student has done this for a few cells, the point is usually made satisfactorily. This step is important because we want the student to set objectives in terms of a type of competence (e.g., collecting information for dissemination and marketing) rather than only in terms of just four specific items (e.g., Nr 49, 50, 51 and 52). After working with the student to think through the task of setting objectives and computing the discrepancies between current averages and objectives for a few cells, students may be left to complete the task on their own. The instructor should review the completed form with each student in order to determine the student's understanding of and reasoning for the objectives that have been set. Obviously, if there is evidence of unrealistic expectations, the student should be encouraged to modify the objectives to bring them more in line with what the student can attain.

In terms of the example given in Figure 6, this student initially aimed for 5.0 averages for the objectives in all three DD&E contexts. However, after considering the very low current average in Dissemination and Marketing, several cells in this column were reevaluated, and the instructor and student mutually agreed on objectives with a slightly lower average in this context area (4.8).

---

*Actually, this is a fairly ambitious overall objective. In many cases an average closer to 4.5 may be more reasonable. Because this student had some prior R&D work experience, had already attained "5" and "6" ratings in a number of items, and strongly wanted to work in the educational R&D field, the objectives appeared attainable.

In terms of skill areas the student especially wanted to work on planning skills, so despite a current average rating of only 2.5, the student set objectives in this row which called for major attainment in all three DD&E cells. The areas with largest discrepancies between current and desired competence levels are circled. Major discrepancies exist in all of the Dissemination and Marketing cells and in all of the Planning cells. In addition there is a large discrepancy in the cell Communicating in the context of Development. Perhaps notable is the fact that two cells show zero discrepancy Producing and Communicating in the context of Field Test and Evaluation. Overall, these particular discrepancies point to the need for a fairly comprehensive program in all facets of DD&E, but clearly with the need for practical work experience of some kind to accompany the academic phase in order to acquire practice in a wide range of DD&E skills, and with substantial effort directed to acquiring competencies in the Dissemination and Marketing context and in the Planning and the Analyzing skills areas.

## III. PROGRAM PROGRESS ASSESSMENT

We assume that the student and instructor will develop a tentative plan for a program of instruction based on the assessment and planning at program entry. If the program is truly competence-based, there should be provisions for credit by examination, and advancement should be based on demonstration of attained competence rather than simply modules read or courses taken. All of the DD&E instructional modules specify instructional objectives, and the majority of them contain end-of-module written examinations. Hence, the DD&E instructional materials can provide some assistance in making program progress assessment; however, this information will be primarily in terms of the student's knowledge of technical terms, concepts, and principles. Wherever possible, the instructor should attempt to add information about actual performance (in terms of instructor, peer or work supervisor's ratings, observed performance in laboratory, intern, or on-the-job settings, and in terms of products, such as plans, outlines, analyses, prototype products, evaluation reports, etc., which the student produces). Our field tests of various versions of an assessment battery consisting of (a) self-ratings, (b) knowledge tests, and (c) performance test simulations, which are discussed in section IV, indicate that while each of these three sources of information have some validity, they are far from redundant. Moreover, knowledge tests appear to be the weakest of the three sources of information, at least in terms of discriminating between students and professionals. Appraisals of students' products and ratings of specific competence areas made by those who have been able to observe the student's behavior in work situations (e.g., instructor, peers, or work supervisors in laboratory or real work settings) can add important information, which should be considered in assessing competence levels.

We are able to offer the instructor only modest assistance in this regard. With respect to the appraisal of products, we attempted, and failed, to produce a "general" product rating form that could be employed with a variety of products. However, we have been able to develop a scoring manual for the simulation tasks described in section IV (see also Appendix E ). Examination of the methods of scoring used in this manual may provide ideas for scoring other products. Basically we have attempted to appraise each product in terms of four types of questions:

1. (Pass/Fail) Is the product acceptable? This is a pass/fail determination which must be based on the instructor's judgement. In some

instances the instructor may have sufficient experience to be quite sure about the judgment. In many cases, it may be illuminating to ask a few R&D professionals who actually supervise entry-level professionals to examine a sample of student products and make their own appraisals in terms of whether they would accept the product or not, and why. Our own personal experience with critiques of the simulation tasks we have developed indicates that this can be a highly useful kind of "reality testing" which may cause the instructor to refine or even significantly modify the criteria for pass/fail.

2. (Fractional Pass) How close did the product come to passing, and what is unacceptable about it? If the product is not acceptable, answers to these questions are essential if the student is to obtain feedback on performance and take appropriate corrective action. Emphasis on what was done properly can provide positive reinforcement for study and work accomplished. Specification of precisely what aspects require further attention gives the student needed guidance on how to produce a better product.

3. (Scope of Application) In what way does this product, regardless of pass/fail on acceptability, reflect application of relevant knowledge, skill or sensitivity that goes beyond the particular minimum requirements for acceptability? We have attempted to score products for "scope" or "breadth" of application because we have encountered substantial variance in the overall quality of the products produced in response to the specific requirements. Criterion referenced testing is often presented as a "pass/fail". Our second appraisal question addresses the problems of degrees of failure. This question addresses another problem, the overall quality of the performance. Feedback to the student here is not in terms of what must be corrected to produce a passable product, but rather it is on identification of and positive reinforcement for any outstanding feature of the product.

4. (Specific Component Scores) What can be said about the quality of specific aspects of this particular type of product? While the previous three questions reflect a global appraisal of the overall products, most products can be divided into several components which can be apprasied individually (e.g., quality of writing, evidence of use of appropriate references, conformance with specifications). Our scoring in this situation is related to the "fractional pass" score discussed under #2 above; however, for the sake of scoring economy we have not attempted to make a close link. The "fractional" assessment deals exclusively with the entire minimum set of elements or characteristics which should be considered in establishing a pass/fail criterion. The "specific components" assessment selects only the most important or major elements of a particular kind of product and singles them out for special attention.

Now these are quite general guidelines. The reader should refer to Appendix E for examples of their application in specific instances.

With respect to performance ratings, we have employed the simple device of changing the person reference of our rating scale (see p.19) from "I" to "She/He," and then asked work supervisors, student peers, work colleagues, or instructors who were knowledgeable of the student's performance to rate the student using items from the DD&E Assessment Instrument or the supplementary item pool. Raters should be encouraged to rate the student on all competencies where they have enough information to feel qualified to make a rating, but should be allowed to omit competence items where they feel that they lack relevant information. (An important discovery for us in this regard was the fact that many immediate supervisors of DD&E professionals were unable to rate the persons they supervised on many of the diagnostic items. In some cases this was because the particular kind of work had never been performed under their supervision, but in many cases it was because they simply did not supervise closely enough to have any good impression of the employee's capability.)

On a more general note, performance ratings, if supplied by others, and most certainly if supplied by peers, should be attempted only in a "formative evaluation" context and with provision for full and open exchange between rater and ratee. Under these conditions, the ratings can serve to establish positive and beneficial communication. Anonymous ratings by others can be highly useful in one-shot, go/no-go assessment, but are not recommended in this instance, because the student needs to "verify" the basis for the rating, and may need help in doing something about it if it is below his/her expectation.

In Section II we outlined the type of file which needs to be kept on student progress. In that section we also introduced the DD&E Diagnostic Instrument Record Form, and illustrated how it could be used as a device to initiate program planning by recording (a) entry level competence ratings, (b) program objectives ratings, and (c) the discrepancies between these two ratings for each cell in Figure 6. It seems to us that a simplified extension of the device could be used to provide a summary of the student's accomplishment in each area. One possibility would be to use a simpler matrix form (such as Figure 4) and enter in each empty cell of that matrix pertinent data representing progressive estimates of competence levels, as well as revised statements of objectives. For a possible example of a series of cell entries see Figure 7.

Figure 7

SAMPLE ENTRIES IN A STUDENT PROGRESS RECORD MATRIX

| | Development | | F.T.&Eval. | | Diss.&Mkt. | |
|---|---|---|---|---|---|---|
| Collecting Information | 4.0 | 5.0 | 4.5 | 5.5 | 2.5 | 5.0 |
| | 4.3 | --- | 5.0 | --- | 3.5 | --- |
| | 4.4 | --- | --- | --- | --- | 4.5 |
| | 4.8 | --- | 5.5 | ✓ | 3.8 | 4.2 |
| | 5.2 | ✓ | | | 4.5(+.3) | ✓ |
| Analyzing | 4.0 | 5.0 | 3.0 | 5.0 | 1.5 | 4.0 |
| | 4.2 | --- | 3.6 | --- | 3.0 | --- |
| | --- | 5.3 | 3.7 | --- | --- | --- |
| | 5.0 | --- | 4.2 | 4.5 | 3.8 | 4.0 |
| | 5.1 | (-2) | 4.6(+.1) | ✓ | 4.0 | ✓ |
| | | | | | | |

line 1 -- 10/12/76
line 2 -- 12/3/76
line 3 -- 1/17/77
line 4 -- 3/16/77
line 5 -- 5/11/77

In each cell assessed level of competence, as agreed upon by student and instructor, appears on the left and current objective appears on the right. Date of entries of each line are recorded on the bottom of the form. Dashes indicate no change from previous agreed upon levels for actual attainment or for objectives. Check marks (perhaps in color) signal attainment of the objective. Entries such as (+.1) indicate where final level exceeded the objective. Note the final entry in analyzing in development is (-.2), indicating that the student fell short of attaining the objective in this area.

Such a summary status sheet on estimated competencies and objectives would be relatively easy to maintain and can provide at a glance the overall picture of the student's progress. Obviously the data contained in such a form can be but a numerical "shorthand" for the much richer set of notes, records, reports, and other materials which must be contained in each student's file.

As a final note, our assumption is that, despite their brevity, the competence rating levels illustrated in these summary status sheet entries, will be based on a joint assessment made by the student and instructor, taking into account all of the available evidence at each point in time. Hence, these are not simply self ratings, or instructor ratings — but rather an overall assessment based on end of module test scores, performance on student projects, appraisal of reports and papers, and other kinds of evidence that may bear on making a reasonable inference about the student's current level of proficiency in each major competence area.

## IV. EXIT COMPETENCE CERTIFICATION

### A. ASSESSMENT OF EXIT COMPETENCE

It seems desirable for a competence-based training program to have some systematic means for reviewing evidence of attained competence and for certifying student attainment at the time the program is completed. This can be done by reviewing the entire student file and making a final set of appraisals, which should include not only a final set of proficiency ratings, but summary annotations, based on the instructor's notes and other documentation, regarding the character and quality of the student's performance. This type of documentation, will provide the student, the instructor and prospective employers with a useful form of competence-based certification.

However, we felt there was a need for some type of final examination, that might be used at the end of an instructional series or at the end of the entire DD&E program.

After pilot development and test of two "mini-batteries," each based on a single DD&E instructional module, it was apparent that the project would not have the time and resources to develop and validate certification materials for the entire DD&E program. We narrowed our efforts to the development of a Competence Assessment Battery (CAB) that would certify skills in the area of "product development." The DD&E Training Program has five Development Series instructional modules, which cover product design, production engineering, review, tryout, and revision. The five modules provide *orientation* and *familiarization* with the general procedures and problems employed in product development. Because review, tryout, and revision are considered in the series, the instructional content of the series spans both the *development* and the *field test and evaluation* contexts of the DD&E competence matrix (Figures 4 and 5). Moreover, virtually all of the process skills can be encompassed. For these reasons, this series appeared to be the most promising both in terms of validation of a Competence Assessment Battery (CAB) "model," and in terms of providing DD&E program users with a core set of CAB items.* The final version of the Development Series CAB consists of three items:

- RDD&E Activities (a rating form)
- Job Knowledge Questions (a short answer essay knowledge test)
- An Exercise in Educational Development (a performance simulation)

## 1. RDD&E Activities

This rating instrument consists of 60 items, similar in character to those appearing in the DD&E Diagnostic Instrument, but this time including competence content dealing with research and statistics, planning and writing, and management, as well as development, evaluation, and dissemination.**

The RDD&E Activities items can be self-rated by the student. And in some cases the student may be rated on the same items by others (e.g., instructor or work supervisor). The instrument is self-administering and usually requires ten to fifteen minutes to complete. The instrument provides scores on six 10-item scales: (1) Research and Statistics, (2) Evaluation, (3) Instructional Product Development, (4) Planning and Technical Writing, (5) Dissemination, and (6) R&D Management.

---

*In our opinion, the next most useful addition would be to develop CAB items in the Dissemination and Marketing Context in order to provide some coverage of all three contexts.

** In developing the DD&E instructional materials, we assumed that most students would also take courses in research and statistics to augment their work in DD&E. Although rarely taught as a specific course, the AERA data had sensitized us to the existence of R&D management competencies. Hence, it seemed desirable to include a sampling of items in these other content areas.

The validation field test version of the self-rating instrument employed a simplified six level performance scale which omits all references to knowledge and focuses solely on performance.* The scale is as follows:

### Level of Performance

1. I cannot perform this activity even with supervision or guidance.
2. I can perform this activity if I have either detailed instruction or close supervision.
3. I can perform this activity if I am given enough general supervision or general instructions.
4. I can perform this activity quite satisfactorily without supervision or job aids.
5. I can perform this activity quickly and efficiently and can do a top quality job.
6. I consider myself an expert in this activity and can accomplish unusually difficult or completely novel work.

---

*Please compare this performance scale with the proficiency scale used with DD&E Diagnostic Instrument (p.19). The first three levels of the proficiency scale (1, no knowledge; 2, don't really understand; 3, good understanding but haven't done it) have been combined into the rough equivalent of performance level 1 (can not perform this activity even with supervision or guidance). Proficiency scale level 4 corresponds approximately to performance scale level 2; 5 corresponds to 3; 6 corresponds to 4, and 7 corresponds to 5. Finally a sixth performance level ("expert") was added. A rough but probably satisfactory comparison between the scales can be made by adding two points to the performance scale or by subtracting two points from the proficiency scale.

## 2. Job Knowledge Questions

This short-answer essay knowledge test consists of ten items dealing primarily with instructional product development and evaluation. This test had a time limit of one hour. The ten questions involved the following content.*

(1) Identify important information missing in a product description.

(2) Specify appropriate terminal objectives for a briefly described product.

(3) Analyze and discuss a proposed product tryout in terms of the relevance and importance of the information it would yield.

(4) List three of the basic criteria for judging the adequacy of revisions which have been made in [any] product.

(5) Compare the advantages and disadvantages of sound movies and sound filmstrips as instructional media considered from the viewpoint of the developer, the distributor, and the user.

(6) Indicate three questions a developer might want to answer in a tryout about student satisfaction with a programmed text.

(7) Identify the information that could be obtained from a consultant review of a product that might not be obtained in a field test of the product.

(8) List three questions (for each of three specified types of reviewers) which should be asked about a [specified] product.

(9) Identify factors other than the product itself, which might account for difficulty subjects may encounter in using [any] product effectively.

(10) Outline the sequence of major steps that are most important in the production of a sound filmstrip between the two stages (a) definition of the instructional objectives and (b) collection of the first pilot test data.

---

*The content listed here is sometimes abbreviated — see Appendix D which contains the complete form of each question and the scoring instructions.

Each item was scored three ways: (a) pass-fail (P-F), scored 1 or 2; (b) fraction (F) of essential elements covered in the response, scored 0 - 9; and (c) scope (SC) of application of knowledge going beyond essential requirements, scored 1 - 5.

### 3. An Exercise in Educational Development

This is a self-administered simulation test which requires approximately four hours to complete. The simulation contains an introduction which explains how the simulation is organized, what the student is expected to do, defines the student's role as a member of the simulation's development team, and provides background information regarding the product. The team is working on a manual for training librarians of a "toy library" how to teach parents to use educational toys with their children. The introduction is followed by four sequentially organized tasks:

Task 1. Preparing Guidelines for Developers of the Librarian's Manual. (1-1/4 hour time allowed) This task requires the test subject to read several documents, extract relevant information, and organize it in a memo so the development team will know for whom the manual is to be written, in what form it is to be prepared, what it is supposed to accomplish, the specific steps the librarian should follow in conducting a parent training session with the selected toy, etc.

Task 2. Preparing for the Tryout of the Resource Kit. (1/2 hour) The subject is required to do two things in this task: (1) describe what the Research and Evaluation (R&E) Team (which would conduct the field testing for the development project) should look for in selecting test sites, librarians, and parents in order to conduct a field test of the librarians manual, and (2) list the questions (separately for librarians, parents, and R&E observers) that the subject would like to have answered in the tryout.

Task 3. Recommending Revisions for the Librarian's Manual. (1 hour) In this task the subject is presented with a five-page summary report of the tryout of the librarian's manual at two test sties. On the basis of this summary the subject is asked to do two things: first and most important, list suggestions for revising the librarian's manual based on the tryout results; second, critique the tryout and recommend what additional instructions should be given to the R&E team before undertaking further tryouts.

Task 4. Outlining a Script. (1 hour) This is a far more specific task that would be appropriate only if the student has developed some competence in development of audio-visual instructional materials. It calls for the subject to rough outline the content of a short sound filmstrip designed to provide a "model" for parents to view before attempting to role-play "parent" and "child" roles in practicing to use the toy with their children. In addition the subject is asked to produce some detail (both audio and visual) for only the first several frames.

Appendix B reports on the development, pilot test and field test validation of the competence assessment battery. Appendices C,D, and E contain the instruments and scoring instructions.

## B. RESULTS OF THE VALIDATION TEST OF THE BATTERY

### 1. Validation Test Subjects

Sixty-six subjects were recruited from ten sites, including four universities and six R&D agencies. In terms of experience the validation subjects may be divided into three groups: 13 students who had completed relevant course-work* but had no educational R&D work experience; 21 subjects (mainly students or recent graduates) with modest work experience and some record of producing R&D publications or products; and 32 subjects with significant R&D work experience and a substantial record of publications and products. Table B1 of Appendix B provides data on sex, average age, teaching experience, R&D work experience, and numbers of publications and products for the three groups. Table 4 summarizes these data.

There are substantial differences among the three groups in age, years of R&D work experience, number of publications and number of products. Although the majority of all three groups have some teaching experience, nearly all of the modest experience group have such experience. Note also that the high R&D experience group is preponderantly male, but that females are in the majority

---

*Typically, these students had completed one course in research design, one course in tests and measures, one course in instructional technology and one or two courses in statistics. About half had also completed a course in evaluation.

# TABLE 4

## BIOGRAPHIC DATA ON THE VALIDATION GROUPS

| | R&D EXPERIENCE AND PRODUCTIVITY | | |
|---|---|---|---|
| | No | Modest | High |
| Number in group | 13 | 21 | 32 |
| Age in years | 24.4 | 30.3 | 38.4 |
| Percent male | 31% | 38% | 84% |
| Percent who have taught | 62% | 91% | 69% |
| Years R&D work experience | 0.2 | 1.4 | 10.1 |
| Number of publications* | 0.3 | 2.2 | 17.6 |
| Number of products* | 0.2 | 5.4 | 21.3 |

*Note that figures for number of publications and number of products are sums of averages for five separate categories of publications and five separate categories of products which are reported in Table B 1 of Appendix B .

for the other two groups. The average individual in the modest R&D experience group has worked in some type of educational R&D work setting for 1.4 years, has produced two publications and five products. The average individual in the high experience group has worked ten years and produced over 17 publications and 21 products.

## 2. Results for the RDD&E Activities Self-ratings

All six scales of this self-rating instrument were significant when tested by one way analysis of variance based on scores of the three R&D experience level groups. Table 5 summarizes the results.

The student group with no R&D work experience (but with some relevant training) attains average scale ratings which are close to or slightly above 3.0 ("can perform if given enough general supervision or general instructions").* By contrast, the high R&D experience group has average scale ratings above 4.0 ("I can perform this activity quite satisfactorily without supervision or job aids") on four scales and above 3.7 on the other two (research & statistics, dissemination). The group with modest R&D experience has scale average ratings that are intermediate on all six scales.

It should be noted that the items appearing on the six self-rating scales have not been selected on the basis of group discrimination; but rather for their coverage of RDD&E content that seemed appropriate for entry level professionals.** Given the very large differences in work experience and productivity indicated in Table 4, the general character of the self-rating results reported in Table 5 should be expected. For instance, note that the high R&D experience group has average ratings that are substantially higher than the other two groups on the R&D Management Scale and on the Dissemination Scale. In these areas, the members of the modest experience group have not rated themselves appreciably higher

---

*Since there are ten items on each scale, the averages reported in Table 5 may be read as performance scale ratings by shifting the decimal one place to the left (i.e., the 30.2 average of the group with no R&D experience on the Research and Statistics scale is 3.02 on the performance rating scale).

**Table B2 of Appendix B presents group means and p-levels of analysis of variance tests of differences among the three groups for all 60 items.

## TABLE 5

## RDD&E ACTIVITY RATING SCALE MEANS
## AND SIGNIFICANCE (P-LEVEL) OF F-TESTS FOR VALIDATION GROUPS

| SCALES | R&D EXPERIENCE LEVEL | | | P-LEVEL |
|---|---|---|---|---|
| | No | Modest | High | |
| 1. Research and Statistics | 30.2 | 33.2 | 38.3 | .05 |
| 2. Evaluation | 31.3 | 34.3 | 41.6 | .01 |
| 3. Instructional Product Development | 33.2 | 38.8 | 41.5 | .01 |
| 4. Planning and Technical Writing | 33.7 | 37.3 | 43.2 | .01 |
| 5. Dissemination | 31.2 | 31.6 | 37.2 | .01 |
| 6. R&D Management | 30.5 | 31.4 | 41.9 | .01 |

than the no experience group. However, on the other four scales, the modest experience group has average ratings that are more nearly intermediate, but closer to the no experience group than the high experience group in Research and Statistics and in Evaluation; and closer to the high experience group in Instructional Product Development and in Technical Writing.

Although our focus in validation is mainly related to work experience criteria, we note that the self-rating scales are also significantly and differentially related courses taken. For instance, there are modest to substantial correlations (.41 to .63) between both the Research and Statistics and the Evaluation scales and the number of courses taken in Research Design, in Tests and Measures, and in Statistics, as well as in "other" (than educational) Psychology courses. Hence there is evidence that the self-ratings reflect the individual's level of training and work experience.

Scale means, standard deviations, alpha reliability coefficients, and correlations are reported in Table B3 of Appendix B. That table indicates that all six scales have internal consistency reliabilities of .90 or higher and that the intercorrelations among the scales range from .40 to .82. The Research and Statistics scale is strongly correlated (.79) with the Evaluation scale, but otherwise displays the lowest correlations with the remaining four scales (.40 to .51). Instructional Product Development is the next most independent scale with correlations ranging from .51 (with Research and Statistics) to .74 (with Planning and Technical Writing). By contrast, the Evaluation scale shows moderately high correlations with all five other scales (ranging from .66 to .79). The Management scale also displays moderately strong correlations with all scales; the highest being .82 with Dissemination and the lowest being .51 with Research and Statistics.

To summarize, the field validation data indicate that despite their brevity (ten items) each of the six scales has high internal consistency, is able to discriminate among subjects with different degrees of work experience, and tends to correlate with number of relevant courses taken. The size and range of the scale intercorrelations (.40 to .82) indicates that individuals tend to rate themselves at roughly similar levels on all six scales, but that the scales are sufficiently independent to treat each as a separate scale.

### 3. Results for the Job Knowledge Questions

Each of the ten essay questions in this test were scored three ways: pass-fail (scored 1 or 2); fraction of essential elements covered (scored 0 to 9); and scope of application of knowledge going beyond essential requirements (scored 1 to 5). Because the pass-fail (P-F) score is operationally related to the fraction (F) score, it is not surprising that the correlation between these two scores, averaged over the ten items, is .87. The scope (SC) score, which may be considered a "bonus" score for responses which reveal a "scope" of application of knowledge going beyond the specific essential requirements set for each question displays substantially lower average correlations (.48 with fraction and .44 with pass-fail), indicating that the scope score does contain information not present in the other two scores. Although there are significant intercorrelations among these three scores for each item, the intercorrelations of scores among the ten questions were surprisingly small, with only .25 (significant at $P < .05$). A factor analysis of the 20 scores produced as many factors as there were questions. Apparently the knowledge required to deal with these ten Job Knowledge Questions is moderately specific to each item; however the questions are not totally independent. For instance, there are very modest but significant correlations among the several questions dealing with review of tryout, and there is a tendency for fraction or scope scores of one question to be related to scope scores of other questions, but none of these relationships are particularly strong.

Analysis of variance of score differences among the three experience groups indicated that only four of the ten questions had score mean differences that were significant at the .05 level. However the total scores (summed over the 10 items) are all significant. See Table 6.

Note that the modest R&D experience group has slightly higher total scores than the high experience group (all non-significant). Both of these experienced groups have significantly higher scores than the no experience group. Expressed as percent of the possible score we see that approximately 15% more of the experienced subjects "pass;" however, the absolute percentages passing are troubling, primarily because they indicate that even the more experienced groups were averaging only 6.9 and 6.7 questions passed out of ten. Although the one

## TABLE 6

## MEANS, F-RATIOS AND P-LEVELS FOR THE F, P-F, AND SC TOTAL SCORES ON JOB KNOWLEDGE QUESTIONS

| | R&D EXPERIENCE & P-LEVEL | | | F-Ratio | P-Level |
|---|---|---|---|---|---|
| | No | Modest | High | | |
| Pass-Fail (P-F)<br>(P-F as % possible) | 15.3<br>(53%) | 16.9<br>(69%) | 16.7<br>(67%) | 3.81 | 0.027 |
| Fraction (F)<br>(F as % possible) | 70.6<br>(78%) | 79.5<br>(88%) | 77.9<br>(87%) | 4.89 | 0.011 |
| Scope (SC)<br>(SC as % possible) | 25.2<br>(50%) | 31.7<br>(63%) | 29.8<br>(60%) | 7.54 | 0.002 |

hour time limit may be partly the problem, it seems, in retrospect, that the pass-fail scoring standards (see Appendix D ) were unrealistically high. With a few exceptions, a "pass" score is obtained only if the subject's response included all of several "essential" elements. If even one element was missing the response was given a "fail" score. We would recommend that the user adapt this test by both reducing the number of questions and by relaxing the pass-fail criteria.

The fraction scores are based on a 0-9 scale for each item, which yields a possible range of 0 to 90 for the totals. Expressed in terms of percentages of the possible score of 90 (in parentheses), the fraction scores for students with no R&D work experience averaged 78% of the possible while the more experienced subjects averaged approximately ten points higher. The scope scores also indicate that there are marked differences between the no experience and the two experienced groups. In this case the F-test indicates that the difference is highly significant.

Perhaps because the Job Knowledge Questions are fairly specific, there are few significant correlations between this test and number of courses. All three totals correlate significantly (.27 to .30) with number of courses in Instructional Technology. The only other significant correlations are .243 for pass-fail and number of Creative Writing courses and .26 for scope and number of courses in Philosophy and Humanities.

To summarize, the Job Knowledge Questions test does discriminate between those with no R&D work experience and those with some experience. The differences are not large, but they are statistically significant. However, there is no evidence that those with substantially more experience perform better on this test. In fact, the scores of the high experience group are slightly lower than those of the modest experience group. The three total scores have substantial intercorrelations (.88 between F and P-F; .73 between F and SC; and .71 between P-F and SC). Intercorrelation of scores based on each item indicates that the ten items are not highly correlated, indeed only 16% of the between question score intercorrelations were significant and none exceeded .50. Individual items generally did not show strong differences between the experienced and inexperienced groups, but the total scores are all significant.

Given the results, we would caution the user against generalizing on the basis of a student's response to only one or two questions of the type used in this test. However, responses to several questions apparently yield scores which have credible validity.

Because of the relative specificity of individual items, the instructor should exercise care in selecting or writing items that broadly sample knowledge in the competence areas of interest. The particular items used in this test were deliberately written to roughly parallel the content of the simulation test, hence they encompass a decidedly narrower range of knowledge than the competencies sampled in the self-rating instrument. Although the items might be satisfactory for use in testing exit competence in the DD&E Development Series or in a comparable course in instructional product development, they are far too narrow in scope to cover the entire DD&E Program as suggested in Figures 1 and 5.

### 4. Results for the Exercise in Educational Development

Like the Job Knowledge Test, each of the four tasks of this simulation were scored for pass-fail (P-F), fraction (F) and scope (SC). Each task was also scored in terms of the extent to which the simulation directions (DIR) were followed. In addition, the four tasks were scored on a total of eleven specific elements (e.g., the quality of writing of the guideline produced in Task 1; the quality of the tryout questions listed in Task 2, the suggested revisions in Task 3, the handling of the visual elements in the script outline in Task 4).

The sixteen common scores (P-F, F, SC and DIR) and the eleven specific scores were correlated and factor analyzed. These results are reported in Appendix B. Summarized briefly, there were moderate to substantial intercorrelations among the several scores based on the same task, but smaller correlations among scores for different tasks. However, in contrast to the small proportion of significant intercorrelations found in the knowledge test, fully one half of the between task correlations were significant. A factor analysis indicated that the first principal axis accounted for 34% of the trace; in other words approximately one third of the covariance among the scores was attributable to a general performance factor. All of the 27 scores displayed significant loading on the general performance factor but these loadings ranged from .33 to .81. Hence the several scores exhibit substantially different degrees of association with this general performance factor.

When the six factors were rotated to a varimax solution, we found that Task 1 (preparation of product guidelines) was associated primarily with one factor and Task 4 (outlining a script) was primarily associated with another factor. However the scores in Task 2 (preparing for the tryout) and Task 3 (recommending revisions based on tryout results) loaded on several different factors. We conclude that the simulation represents a complex of several different performance dimensions.

When we examined individual item scores for differences among the three experience groups, we found that the fraction and the scope scores, as well as several of the specific scores for Tasks 1, 2 and 3 had F-ratios that were significant; however, none of the scores associated with Task 4 reached the required significance levels.

Five total scores were available. In addition to pass-fail, fraction, scope, and following directions, we obtained a consistency total score which is the sum of four specific task scores; one from each task which best reflects the test subject's ability to cope with and respond consistently with the information (not the directions) provided in the simulation. Table 7 reports the results for these five total scores.

As in the case of the Knowledge Test, there is a problem with overly stringent pass-fail criteria; neither of the experienced groups has attained even 50% of the possible score. Nevertheless, both of the experienced groups perform significantly better than the no experience group. The fraction score is more informative. It indicates that 72% of the possible fraction total score was attained by the no experience group, while the modest experience group attained 84% and the high experience group 85%. Scope performance averaged lower (52%, 70% and 70% respectively), but the difference between the no experience and the two experienced groups is larger and highly significant ($P < .001$). All three of the groups tended to perform well on following directions, but even

## TABLE 7

### Means for Three R&E Experience Groups, F-Ratios and P-Levels for Tests of Differences Among the Groups on Simulation Test Scores

| | R&D EXPERIENCE LEVEL | | | | |
|---|---|---|---|---|---|
| | No | Modest | High | F-Ratio | P-Level |
| Pass Fail<br>(P-F as % possible) | 4.38<br>(10%) | 5.71<br>(43%) | 5.31<br>(33%) | 5.38 | .007 |
| Fraction<br>(F as % possible) | 25.77<br>(72%) | 30.10<br>(84%) | 30.56<br>(85%) | 5.90 | .005 |
| Scope<br>(SC as % possible) | 10.38<br>(52%) | 13.95<br>(70%) | 14.00<br>(70%) | 9.17 | .001 |
| Following Directions<br>(DIR as % possible) | 32.46<br>(90%) | 34.33<br>(95%) | 34.72<br>(96%) | 3.17 | .050 |
| Consistency with Sumulation Information (% possible) | 23.92<br>(66%) | 29.67<br>(82%) | 28.56<br>(79%) | 7.03 | .002 |

here there is a significant difference ($P < .05$). Finally, both of the experienced groups were better able to cope with the mass of simulation information than the no experience group.

## 5. Relationship Between Self-Ratings, Knowledge and Performance

Table B7 of Appendix B reports the intercorrelations among the scale totals for the three types of instruments. These findings may be briefly summarized as follows. Knowledge as measured by all three scores of the Job Knowledge Questions is modestly, but significantly, related to all five of the simulation test total scores, with correlations ranging from .25 to .46. Given the fact that the Job Knowledge Questions were written to cover general content, roughly paralleling the content of the simulation tasks, the fact that every one of the fifteen correlations between these two instruments is significant is not surprising. Since the Job Knowledge Questions tend to cover relatively general types of knowledge and are clearly not concerned with testing for strictly "enabling" knowledge that would be specifically needed to complete the simulation tasks satisfactorily, perhaps the modest level of observed correlations (.25 to .46; average .38) should be expected. These data do indicate that, unless knowledge content is clearly "enabling," one should not expect to find high correlations between knowledge and performance on simulation tasks. The average correlation of .38 indicates that there is less than 15% of the total variance of a performance score which is predictable (or explained) from a knowledge test score. This is hardly a strong relation despite the fact that it is statistically significant from zero correlation.

One or more of the knowledge test total scores is also significantly related to each of the three self-rating scales which most closely corresponds to the Job Knowledge Questions content, namely the Product Development, the Evaluation, and the Technical Writing self-rating scales. However, these correlations are also of quite modest size. The two largest ones are .38 for knowledge fraction score with the Product Development self-rating scale; and .41 for the knowledge scope score with the same self-rating scale. None of the correlations between the three knowledge test total scores and the remaining three self-rating scales (Research, Dissemination, and R&D Management) are significant.

Hence the knowl-dge test does display convergent and discriminant validity in terms of its correlational pattern with the self-rating scales.

Although the significant relationships are extremely modest, there is clear evidence that relevant, general knowledge as measured on responses to short answer essay questions is related both to self-ratings and to performance on simulation test.

However, there are no significant correlations between the self-rating scales and the simulation test total scores. This result was not anticipated, since there were significant correlations between these two types of instrument in field tests of earlier versions of the battery (see Appendix B, pp. B-2 and B-7). The reasons for this result are discussed in some detail in Appendix B, pp. B-32 to B-34. Stated briefly, the correlations with biographic data (numbers of courses taken, products and publications authored) suggest that the three CAB instruments represent increasing degrees of competence specificity with the self-ratings the most general, the knowledge test intermediate, and the simulation test the most specific. Our guess is that the items of the RDD&E [illegible] self-rating scale may be too general in their scope relative to the particular performances which are assessed in the simulation test, and hence, there is no strong relationship despite the superficial similarity of content in the two types of instruments.

## C. CONCLUSIONS

Possibly the most important practical points to be made are these:

All three instruments are valid, both in terms of content validity and in terms of their demonstrated ability to discriminate between subjects with and without actual product development and evaluation experience.

Even where there are significant correlations between types of instruments, the relationships are very modest; hence each instrument tends to provide a different kind of valid information about competence. Since most professional training programs seem to rely heavily on paper and pencil tests of knowledge, it may be comforting to know that significant correlations were found between the essay test scores and both the self-ratings and the simulations. However, the correlations are so small that one needs to question the utility of an assessment system that relies exclusively or almost exclusively on measurement of knowledge.*

Because the several types of assessment instruments are reasonably independent of one another, and yet all are valid, their combined use should increase the overall validity of competence assessment decisions. (See Appendix B, footnote on p. B-35 for one example.) Moreover, inclusion of other types of information (see pp. 12 and 14) is recommended.

Competence assessment is obviously a "sampling" proposition. Our field test results suggest that there are problems which deserve special attention. Given the generally low correlations among items within knowledge test and within the simulation test, one must exercise caution in making generalizations except where the number and scope of test items are sufficient to adequately represent the competence area being assessed. The high degree of specificity of each task in the simulation test, and its factorial complexity are indeed troubling. Our concern is that assessment based on performance on one or only a few simulation tasks may lack substantial general validity. Apparently a simulation task is analogous to a single test item in terms of reliability and validity; we need a sample of several tasks to achieve useful reliability and validity.** The four tasks comprising the current simulation are almost too much to handle at one test session, and yet this amount of simulation performance information seems hardly sufficient. Several more tasks are probably required to provide even minimally satisfying samplings of even the educational product development competence domain. Given the substantial costs in student and instructor time that need to be invested in this kind of performance assessment, we doubt that simulations will be attractive unless they are built into the instructional program and treated as significant learning opportunities (with adequate instructor feedback and counseling) rather than as solely assessment exercises. Please note that we believe the simulation tasks should be a part of a competence based program. Many of the students who took the simulation test indicated that it was one of the few times that they were actually challenged to see what they could "do" rather than just what they "knew" about educational product development. And, despite its artificiality in terms of time limits, lack of team interaction, inability to ask questions or seek assistance, most R&D professionals considered the simulation to be a valid and meaningful challenge.

---

*It is quite conceivable that different results might be obtained if one were to employ a more comprehensive multiple-choice type test. However, based on our very early work with this type of instrument, we doubt that the correlations would be much higher. On the other hand, we suspect that test of specifically "enabling" knowledge would demonstrate substantially higher correlations with simulation tasks or highly specific rating items with essentially the same competence content.

**Note, however that several different elements may be scored on each task, hence a simulation consisting of several tasks may yield several different total scores.

Finally, we must realize that what is presented here is a deliberately "open" assessment system. The assessment model and the resources illustrate what can be done. For those who wish to establish a DD&E training program based on the Far West or similar training materials, the instruments and guides contained in the following appendices will provide a good foundation. But there is a clear challenge to make significant additions or modifications to the resources we have been able to provide.

# Appendix A

## A Report on the Field Test and Revision of the DD&E Diagnostic Instrument

Paul D. Hood
Andrea Lash

Appendix A was submitted as an interim report dated November 1974. It has been reproduced in its entirety here to make it conveniently accessible.

## TABLE OF CONTENTS

SUMMARY. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . i.

INTRODUCTION . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

FIELD TESTING. . . . . . . . . . . . . . . . . . . . . . . . . . . 4

DATA ANALYSIS. . . . . . . . . . . . . . . . . . . . . . . . . . . 5

  Cluster Analysis . . . . . . . . . . . . . . . . . . . . . . . 5
  Levels and Distribution of Proficiency Ratings . . . . . . . . . . 9
  Analysis of Scales . . . . . . . . . . . . . . . . . . . . . . 14

INSTRUMENT REVISIONS . . . . . . . . . . . . . . . . . . . . . . . 20

  Revisions Based on Cluster Analysis. . . . . . . . . . . . . . . 21
  Revisions Based on Item Location . . . . . . . . . . . . . . . . 21

REVISED DIAGNOSTIC INSTRUMENT. . . . . . . . . . . . . . . . . . . 23

APPENDIX A: REVISED EIGHT-POINT RATING SCALE AND ORIGINAL DIAGNOSTIC INSTRUMENT. . . . . . . . . . . . . . . . . . . . . . . . . A-1

APPENDIX B: RESULTS OF THE CLUSTER ANALYSIS . . . . . . . . . . . . B-1

# SUMMARY

The Far West Consortium for DD&E Training has created a set of instructional modules which provide a core of curriculum resources for a competence-based program for the training of entry-level professionals in educational development, dissemination, and evaluation (DD&E). As part of an assessment system, a DD&E Diagnostic Battery was developed to help students and instructors to determine the students' levels of competency in order to plan programs of instruction tailored to individual needs. Because self-ratings have proven to be a viable and inexpensive means of obtaining this competence information, a self-rating instrument, the DD&E Diagnostic Instrument, is the most important part of the Diagnostic Battery. Paper and pencil tests of knowledge and job sample tests constitute the other elements of the battery which may be employed as validity or calibration checks on the self-assessment instrument.

This report describes the field test, data analysis, and revision of the DD&E Diagnostic Instrument. The instrument consists of a sample of 72 competence statements which have been drawn to provide equal representation of the three major work contexts: development, dissemination, and evaluation (24 items each), and simultaneously to provide equal representation of six process skills: collecting information, analyzing, planning and designing, producing, evaluating, and communicating (12 items each, four in each of the three contexts). The student is asked to rate each item on a seven-point scale that combines knowledge and experience to provide a behavioral reference for making judgments about the attained level of proficiency.

The field test of the instrument was conducted at several academic institutions and R&D agencies throughout the country. A total of 78 subjects provided usable data (62 under-graduate and graduate students and 16 developers and senior

developers in R&D agencies). Analysis of the data indicates that the DD&E Diagnostic Instrument items are highly intercorrelated but that at least three or possibly four subscales can be differentiated. Cluster analysis resulted in four well-defined clusters, three associated with the DD&E contexts and a fourth derived primarily from items appearing in the D&D contexts relating to publication, production, and public relations. However, correlation and factor analysis indicate that this last cluster is highly correlated with and exhibits a factor pattern similar to the dissemination cluster. The factor analysis results suggest that only three factor scales are needed: 1) development, 2) field test and evaluation, and 3) production, dissemination, and marketing.

Neither the cluster analysis nor the factor analysis resulted in a group of competence statements associated primarily with any *one* of the process skills; however, both types of analysis indicate that the newly defined DD&E factors exhibit different patterns of skills which appear to be meaningful. The *evaluation factor* is most prominantly associated with: collecting information, analyzing, and evaluating; the *production, dissemination and marketing factor* with: planning and designing, producing, and communicating; the *development factor* with: analyzing, planning and designing, producing and evaluating.

The analysis indicated that the items which had been placed *a priori* in the development *context*, were in fact a mixture of all three DD&E competencies, consequently, the number of items associated purely with development is closer to nine in number rather than the intended 24.

Given the predominately student population of the field test sample (62 of 78 subjects), the average competence level and the distribution of item means seems acceptable, since the average proficiency level for the entire instrument is at the midpoint (4.0) on the seven-point scale, and the great majority of the item proficiency levels are between 3.0 and 5.0 (61 of the 72 items). Analysis of variance between the student and the developer groups indicates that the majority of the items and all of the scales discriminate between these groups. In terms of the proficiency scale, the average difference is 1.27 proficiency levels. The means for the majority of the competency statements for the student group were below 4.0 which corresponds to the statement:

> "I have a general understanding of this activity and have had some experience with it, enough so I can do it if I have either detailed instructions or close supervision."

On the other hand, the majority of the competence statement means for developers were above 5.0, which corresponds to the statement:

> "I have enough experience in performing this activity to do it if given enough general supervision or general instructions."

This clear, meaningful, and highly significant difference ($p<.0001$) provides further evidence that the instrument is able to discriminate between groups with different amounts of relevant experience. (Previous pilot studies have shown that self ratings made by educational developers correlated with supervisors ratings of the same sets of competencies, .73 for one scale and .88 for another, thus indicating that experienced developers and their supervisors exhibit substantial agreement regarding the average levels of proficiency.)

In terms of levels of self-rated proficiency, both the students and the developers reported relatively higher levels of proficiency in development, followed closely by field test and evaluation, and relatively lower levels of proficiency in publication, production, and public relations competencies. Both groups rated communication skills highest and planning and designing skills lowest. The difference between the two groups was at least one proficiency level for all of the subscales (contexts, skills, cluster analysis scales), with the greatest differences appearing on the development context scale, the field test and evaluation cluster scale and on the analyzing, the collecting information, and the evaluating skills scales. The smallest differences between the two groups were on the dissemination and marketing context scale, the (related) publication, production, and public relations cluster scale, and the communication skills scale. In the case of communication, the relatively small difference in proficiency level seems to be attributable primarily to a (relatively) high level of proficiency in the student group with respect to oral and written communication. In the case of the dissemination and marketing context scale and the publication, production, and marketing cluster scale the smaller difference between groups is primarily attributable to the relatively lower levels of proficiency reported by the developers.

Given the results of the data analysis, it was decided to retain the content structure of the instrument (matrix of three contexts by six process skills), but to revise a number of items so that the three DD&E factors would be more clearly and evenly represented and more closely associated with the three context columns of the response form. In addition, an eighth level was added to the proficiency rating scale to separate the expert from the highly experienced. Appendices to this report include copies of both the original and the revised instrument.

Field testing of this revised instrument will be accomplished in connection with the field test of the DD&E Assessment Battery. This will permit correlation of self-reporting information with knowledge test scores and scores on a series of DD&E job samples.

# INTRODUCTION

The DD&E Diagnostic Instrument is a guidance device to be used in planning a student's program of study in conjunction with the use of the DD&E competence based learning resources.[1] The content of the instrument is based on the DD&E Competence Matrix (Figure 1), which structures the important activities in development, dissemination, and evaluation (DD&E) in terms of three contexts (D,D&E) and six process skills. The Diagnostic Instrument is composed of a sample of four items (activity statements) for each cell of the matrix. The task for the student (or other rater) is to rate the 72 items according to a seven-point Proficiency Scale. The Proficiency Scale combines levels of knowledge and experience.

## LEVEL OF PROFICIENCY

1. I have no specific knowledge about this activity nor experience with it.
2. I have read about or seen this activity performed, but have no experience with it and don't really understand it.
3. I have studied this activity or have frequently seen it performed and have a good understanding of it, but I have not yet done it.
4. I have a general understanding of this activity and have some experience with it, enough so that I can do it if I have either detailed instructions or close supervision.
5. I have enough experience in performing this activity to do it if given enough general supervision or general instructions.
6. I have enough knowledge and experience with this activity so that I can perform this task quite satisfactorily without supervision or job aids.
7. I have had extensive experience with this activity, and can perform it quickly, efficiently, and do a top quality job.

---

[1] Readers who are not familiar with the rationale and technical approach for development of this instrument are referred to: Paul D. Hood *et al*., *Development of Assessment Instruments for Competence-Based Education: The Educational Development, Diffusion and Evaluation Training Program* (San Francisco, Ca.: Far West Laboratory for Educational Research and Development, November, 1973).

Figure 1

DD&E Competence Matrix

| Process Skills | Functional Contexts | | |
|---|---|---|---|
| | Development of Educational Products | Field Test and Formal Evaluation | Dissemination and Marketing of Educational Products |
| Collecting Information | I | -- | -- |
| | II | II, II, II, II | -- |
| | -- | -- | III, III, III, III |
| | IV, IV | -- | -- |
| Analyzing | | -- | I |
| | II, II | II, II, II, II | -- |
| | | -- | III, III, III |
| | IV, IV | -- | -- |
| Planning | I, I | -- | I, I |
| | -- | II, II, II | -- |
| | -- | III | III, III |
| | IV, IV | -- | -- |
| Producing and Implementing | I | -- | I, I |
| | -- | II, II, II, II | -- |
| | III | -- | III, III |
| | IV, IV | -- | -- |
| Evaluating | -- | -- | -- |
| | II, II, II | II, II, II | -- |
| | -- | III | III, III, III, III |
| | IV | -- | -- |
| Communicating | I | I | I |
| | -- | II | -- |
| | III, III, III | -- | III, III, III |
| | -- | -- | -- |

Note: Roman numerals tally items by Cluster Analysis (see page 5):

Cluster I - Publication, Production and Public Relations Activities.
Cluster II - Field Testing and Evaluation Activities.
Cluster III - Planning and Evaluating Dissemination Activities
Cluster IV - Development Activities

The instrument is available in two forms: a conventional rating form and a Q-sort deck. In the Q-sort form each of the seven proficiency levels has been separately listed on a 5 x 7 Category Card and each of the 72 activity statements appears on a 4 x 6 Item Card. The users are instructed to place the Category Cards on a table in front of them and then sort the deck of 72 Item Cards by determining their level of knowledge and experience for the item, and then placing the Item Card on the appropriate Category Card. There is no time limit.

A record form, organized in the same form as Figure 1, but with labels, brief statements of the content of each item, and space for recording the scores is provided. When completed, the Diagnostic Instrument record form consists of the rating for each item, along with cell, column, and row summaries, which provide a systematic representation of the student's knowledge and experience in DD&E. The completed instrument is useful to students and advisers for identifying a learner's strengths and weaknesses in DD&E Competencies. It also provides a structure by which student and adviser can discuss DD&E Competencies in relation to the student's particular learning objectives or interests.

The following report describes the field testing and subsequent revisions of the DD&E Diagnostic Instrument.

## FIELD TESTING

The DD&E Diagnostic Instrument is self-administered and takes approximately one hour to complete and record. A list of the items in the Diagnostic Instrument and the rating scale can be found in Appendix A. A total of 78 subjects, representing a wide range of experience and training in educational development, completed the instrument. We have grouped the test subjects into four categories according to their educational background and training. A description of the four categories and the number of subjects in each category is shown below.

| Number | Description of Category |
|---|---|
| 11 | Students recruited from a local university whose backgrounds were similar to that of students we expect to enter the DD&E program. They were seniors or graduate students majoring in education or majoring in the behavioral sciences with an interest in education. |
| 51 | Graduate Students (Masters and Ph.D. candidates) enrolled in educational research and development courses. These subjects were recruited by their professors at Temple University, San Francisco State University, Michigan State University, and UCLA. |
| 12 | Developers - Masters level personnel with at least one year of experience in educational development, or persons with Ph.D.'s in education who have some association with educational development or who are presently developing educational products. Seven of this group were enrolled in inservice training in DD&E. |
| 4 | Senior Developers - Highly experienced in developing educational products and programs. These personnel are presently working in the field. They have published in the field and/or produced products that have been distributed and are being used. |
| 78 | TOTAL |

Please note that this group of subjects is predominantly a student population, which is appropriate since the Diagnostic Instrument is intended for professional education settings. The reader is cautioned to keep this fact in mind while reading the following analysis. The results obtained may not be generalizable to larger groups of experienced R&D professionals.

## DATA ANALYSIS

### Cluster Analysis

A cluster analysis of the data was performed by use of the BC TRY computer package available through the University of California at Berkeley. This data analysis technique groups together variables (items) on the basis of their similarities and differences. It results in clusters of items which are similar to each other and similarly different from other items.[2] The cluster analysis technique was used to determine if the underlying structure of the Diagnostic Instrument (the matrix) would be replicated by the empirical grouping of items.

Four clusters of items resulted. The clusters are most similar to the columns (functional contexts) of the DD&E Matrix. Cluster I is a group of 12 items which describes publication, production, and public relations activities. Cluster II is composed of 25 items which describe field testing and evaluation activities. Cluster III, a dissemination and marketing cluster, is a group of 24 items which describe tasks associated with planning and evaluating dissemination activities. Cluster IV is composed of 9 items describing development activities. Only two items did not appear in a cluster. A list of the items in each cluster can be found in Appendix B. It is noted that none of the clusters were clearly associated with a single process skill (rows in Figure 1). However, the clusters are associated with different groups of process skills. The items in Cluster I involved primarily the process skills of planning, production, and communication. The items in Cluster II involve all six skill areas, but are most frequently represented by analyzing, evaluating, and collecting information skills. The items in Cluster III also involve all six skills, but are most represented by communicating, evaluating, and collecting information skills. The items in Cluster IV are unique in the fact that this is the only cluster in which all items are associated with only one context (development); however, five of the six process skills are represented and in approximately equal proportions.

---

[2] Robert C. Tryon and Daniel E. Bailey, Cluster Analysis (New York, N. Y.: McGraw-Hill Book Co., 1970), pp. 1-5, 45-48.

### *Field Test and Evaluation Competencies*

The cluster analysis results indicate that we can differentiate DD&E competencies. The largest and most clearly defined cluster is composed of field test and evaluation activities. Nineteen of the 25 items in this Cluster (II) are in the field test and evaluation column while the remaining six are in the development column. Four of these six items specifically mention evaluation or field testing. Conversely, only five of the 24 items a priori associated with field testing and evaluation failed to enter Cluster II. On examination, two of these five items (46 and 48) contained no wording that would permit association with any D,D or E context; and these were in fact the two items that did not appear in any cluster in the data analysis.

The first five of the 25 items to enter Cluster II are listed below in the order in which they were added to the cluster (see Appendix B for the entire list):

42. Evaluate test instruments using data collected in tryout and revision cycles in order to recommend instrument revisions for the final field test.

37. Make revisions in test instruments based on evaluation data.

17. Translate field test data into recommendations for action.

39. Prepare a test administration manual.

7. Provided with field test data on instructional materials, examine low gain scores and determine if they indicate problems in test construction or instructional materials.

### *Dissemination and Marketing*

The a priori set of 24 items placed in this column appeared in two clusters (I and III), but 18 of the 24 items were associated with Cluster III: Planning and Evaluating Dissemination Activities. Cluster III also involved four items which had been placed in the Development column and two in the Field Test and Evaluation column. It is notable that four of these six "outside the column" items have "communication" verbs (confer, discuss, interact, and communicate) which suggests that communication, especially informal, oral, two-way communication as well as planning and evaluating dissemination activities are important components of this cluster. The first five items to enter Cluster III in the analysis were:

57. Outline factors which must be considered in disseminating information about an educational product designed for a specific target group.
66. Carry out an evaluation of the effectiveness of a demonstration of and educational product.
59. Plan interviews with potential users for the purpose of determining a market for your product.
55. Review alternatives for the design of a product in terms of possible problems in installing or maintaining the product.
65. Determine the thoroughness of distribution which occurred in dissemination of an educational product.

### *Development*

The items which had been placed a priori in the development column turned out to be a highly mixed lot, with only nine of the 24 actually entering Cluster IV: Development. Four items with "communication" verbs (14, 22, 23, and 24) were associated with Cluster III: Dissemination and Marketing. Six more items were associated with Cluster II: Field Test and Evaluation, and several of these items employed "evaluation" or "field test" references. Finally, five items were associated with Cluster I: Publication, Production and Public Relations. Perhaps this mixed result is only a reflection of our inability to write and place "development" items effectively; however, it may reflect something important about the variety of different skills that are actually associated with development.

The entire list of the nine development items, listed in the order they were added to Cluster IV are as follows:

11. Arrange learning activities in a sequence to facilitate learning or mastery of objectives.

15. Write exercises which the learner should do in order to master concepts or principles in an instructional unit.

19. Provided with product specifications, review a product and documentation on product development and field testing to determine if the specifications have been met.

8. Categorize instructional objectives in terms of a taxonomy (e.g., Bloom's Taxonomy of the cognitive domain or Gagne's conditions of learning).

10. In producing specifications for instructional materials determine an appropriate format for the materials.

16. Write copy for instructional materials from product specifications.

2. Locate existing methods or strategies which can be used in potential product development.

5. Given a problem statement, information on the history of the problem, objectives and possible solutions, write a 20-page proposal for solving the problem, including a rationale for the approach and a development schedule.

3. Prior to conducting a survey of the literature, prepare a search strategy.

### *Publication, Production and Public Relations*

This set of competencies was unanticipated by our original competence matrix, but emerged as a well defined and quite meaningful set of twelve items, consisting of six items from the Dissemination and Marketing column, five items from the Development column and only one item from the Field Test and Evaluation column. The first five items in this cluster are listed below in the order they were added to the cluster analysis:

60. Prepare specifications for audio-visual materials which will be used in the dissemination effort.

61. Write public relations scripts for film and slide shows.

62. Write press releases to disseminate information about a new product.

56. Make arrangements to secure copyrights and copyright clearances where needed.

13. Take steps to assure that such things as waivers, releases, copyright releases or patent protections are secured when appropriate.

## Levels and Distribution of Proficiency Ratings

The average competency rating for the 78 subjects over the 72 items was 4.0 (the median was also 4.0), which corresponds to the rating:

> "I have a general understanding of this activity and have had some experience with it, enough so that I can do it if I have either detailed instructions or close supervision."

Sixty-one of the items were within one proficiency level of this average (3.0 to 5.0). Only five items had average ratings over 5.0 and only six had ratings under 3.0. The average rating and the alpha coefficient of reliability (internal consistency) is shown for the four cluster analysis scales below.

TABLE 1

CLUSTER ANALYSIS SCALES

| | Cluster | No. Items | Average Rating | Alpha |
|---|---|---|---|---|
| I. | Publication, Production, and Public Relations | 12 | 3.6 | .91 |
| II. | Field Test and Evaluation | 25 | 4.1 | .96 |
| III. | Dissemination (Planning and Evaluation) | 24 | 3.9 | .95 |
| IV. | Development | 9 | 4.3 | .90 |

All four of these scales exhibit high internal consistency and satisfactory levels which are neither too high nor too low in terms of average proficiency levels. Examination of the frequency distribution of proficiency levels for individual items did indicate that on nine of the items, over 20 percent of the subjects rated themselves at the 7.0 (highest) level. Interviews with several of the highly experienced subjects indicated that they believed there was a need for an eighth "expert" level of proficiency. An eighth level has been added to the revised scale (See Appendix A.).

*Discrimination Among Groups With Different Amounts of Experience*

As part of the item analysis, each of the 72 items was treated as a dependent variable in a one-way analysis of variance. For this analysis the two groups of students were combined (N=62) and the two groups of developers were combined (N=16). Means for the student group and the developer group, F-ratios ($=t^2$) and p-levels for each item are reported in Table 2. All 72 items exhibit higher means for developers than for students, and for 51 items the significance levels of the differences are .05 or less.

The development context items and the field test and evaluation context items each average approximately 1.4 points higher for developers than for students, with students averaging below 4.0 and developers averaging above 5.0. Only two development items and four field test and evaluation items did not display significant differences ($p < .05$). By contrast, the dissemination and marketing context items exhibited markedly lower averages for both groups and a smaller difference (1.03) between groups. Only nine of the 24 items displayed significant differences ($p < .05$). After reviewing the context of the dissemination and marketing items, it is our conclusion that developers generally aren't as proficient in this area and, therefore, are not as far above students in proficiency as they are in other areas.

TABLE 2

DIAGNOSTIC INSTRUMENT ITEM MEANS AND SIGNIFICANCE TESTS

| ABBREVIATED ITEM CONTENT | MEANS STUDENTS | MEANS DEVELOPERS | F-RATIO | p-LEVEL |
|---|---|---|---|---|
| DEVELOPMENT | (N=62) | (N=16) | | |
| 1. Obtain information on problems | 3.97 | 5.62 | 9.49 | .003 |
| 2. Locate strategies for development | 3.73 | 5.44 | 11.88 | .001 |
| 3. Prepare a search strategy | 4.42 | 5.88 | 8.50 | .005 |
| 4. Evaluate information for relevance | 5.35 | 6.44 | 7.55 | .008 |
| 5. Write a 20 page proposal | 4.06 | 5.00 | 5.04 | .026 |
| 6. Use instructional theories in design | 3.87 | 5.75 | 16.73 | .000 * |
| 7. Review field test data for revision | 3.44 | 4.94 | 8.86 | .004 |
| 8. Classify instructional objectives | 4.03 | 5.50 | 6.59 | .012 |
| 9. Plan a budget | 3.50 | 4.56 | 4.02 | .046 |
| 10. Specify format of materials | 3.68 | 5.31 | 9.40 | .003 |
| 11. Sequence learning activities | 4.37 | 5.19 | 2.70 | .101 NS |
| 12. Estimate needed production materials | 3.81 | 5.69 | 11.59 | .001 |
| 13. Secure waivers, releases, etc. | 2.58 | 3.94 | 6.69 | .011 |
| 14. Confer with specialists when needed | 3.92 | 5.19 | 4.90 | .028 |
| 15. Write instructional exercises | 4.53 | 5.56 | 3.96 | .048 |
| 16. Write copy from specifications | 3.76 | 4.75 | 3.04 | .081 NS |
| 17. Make recommendations from field data | 3.35 | 5.50 | 13.37 | .001 |
| 18. Conduct a case study of a program | 3.95 | 5.00 | 4.45 | .036 |
| 19. Check product against specifications | 3.76 | 5.38 | 10.31 | .002 |
| 20. Informally try out a product | 4.55 | 6.19 | 12.03 | .001 |
| 21. Write a position paper | 3.92 | 5.00 | 4.27 | .040 |
| 22. Discuss a product with user groups | 4.03 | 5.62 | 9.90 | .003 |
| 23. Interact in a staff meeting | 5.21 | 6.38 | 8.41 | .005 |
| 24. Communicate product specifications | 3.58 | 5.62 | 15.14 | .000 |
| TOTAL - DEVELOPMENT | 3.99 | 5.39 | 20.12 | .0001 |

* - All Values tabled .000 have p-Levels of .0005 or less

NS - Not significant at 5% level

TABLE 2 (cont'd.)

| ABBREVIATED ITEM CONTENT | MEANS STUDENTS | MEANS DEVELOPERS | F-RATIO | p-LEVEL |
|---|---|---|---|---|
| FIELD TEST & EVALUATION | (N=62) | (N=16) | | |
| 25. Prepare a coding scheme | 2.94 | 4.19 | 4.45 | .036 |
| 26. Organize statistical data | 4.31 | 5.62 | 7.09 | .009 |
| 27. Interpret scatter plots | 3.65 | 4.81 | 3.98 | .047 |
| 28. Insure privacy in data collection | 3.76 | 5.25 | 9.10 | .004 |
| 29. Discuss internal & external validity | 4.10 | 5.06 | 4.17 | .042 |
| 30. Formulate evaluation questions | 3.98 | 5.19 | 7.63 | .007 |
| 31. Analyze test outcomes | 3.85 | 5.50 | 13.94 | .001 |
| 32. Decide if tests fit evaluation plan | 3.42 | 5.31 | 21.05 | .000 |
| 33. Determine validity of your test | 3.13 | 4.56 | 10.98 | .002 |
| 34. Design a project monitoring system | 2.56 | 4.88 | 21.96 | .000 |
| 35. Plan control of extraneous variables | 3.92 | 4.69 | 2.62 | .106 NS |
| 36. Set criteria for field test sites | 3.37 | 4.94 | 9.99 | .003 |
| 37. Use evaluation data to revise tests | 3.95 | 5.69 | 16.16 | .000 |
| 38. Adjust test procedures when needed | 3.48 | 5.00 | 9.20 | .004 |
| 39. Prepare a test administration manual | 3.15 | 5.06 | 15.82 | .000 |
| 40. Discuss standardized interviews | 4.44 | 5.62 | 6.42 | .013 |
| 41. Decide if statistics are suitable | 3.50 | 4.44 | 3.91 | .049 |
| 42. Evaluate test instruments | 3.60 | 5.56 | 16.04 | .000 |
| 43. Find internal consistency of a test | 3.56 | 4.69 | 5.63 | .019 |
| 44. Identify an evaluation's purpose | 3.76 | 5.62 | 14.96 | .000 |
| 45. Prepare an article for publication | 4.15 | 5.06 | 3.35 | .068 NS |
| 46. Give a short speech or oral report | 6.02 | 6.38 | 1.24 | .269 NS |
| 47. Prepare simple evaluation reports | 4.87 | 6.25 | 11.21 | .002 |
| 48. Prepare graphs to display data | 5.08 | 5.88 | 2.77 | .096 NS |
| TOTAL - EVALUATION | 3.86 | 5.22 | 19.59 | .0001 |

* - All Values tabled .000 have p-Levels of .0005 or less

NS - Not significant at 5% level

TABLE 2 (cont'd.)

| ABBREVIATED ITEM CONTENT | MEANS STUDENTS | MEANS DEVELOPERS | F-RATIO | p-LEVEL |
|---|---|---|---|---|
| DISSEMINATION & MARKETING | (N=62) | (N=16) | | |
| 49. Collect data on installation costs | 2.95 | 4.25 | 5.74 | .018 |
| 50. Design marketing study data forms | 2.79 | 3.69 | 3.75 | .054 NS |
| 51. Use information services | 3.69 | 5.81 | 16.22 | .000 |
| 52. Construct an annotated bibliography | 4.55 | 5.12 | 1.05 | .310 NS |
| 53. Identify character of target group | 3.76 | 4.62 | 2.76 | .097 NS |
| 54. Evaluate market research techniques | 2.34 | 3.69 | 8.17 | .006 |
| 55. Determine problems in installation | 3.08 | 5.00 | 15.18 | .000 |
| 56. Secure copyright clearances | 2.65 | 3.75 | 4.05 | .045 |
| 57. List important dissemination factors | 3.53 | 4.25 | 1.71 | .191 NS |
| 58. Design a public relations activity | 2.71 | 3.50 | 2.18 | .140 NS |
| 59. Determine a market for your product | 3.34 | 4.31 | 3.26 | .072 NS |
| 60. Prepare specifications for A/V material | 3.02 | 3.75 | 2.12 | .146 NS |
| 61. Write scripts for film & slide shows | 3.19 | 3.94 | 1.84 | .176 NS |
| 62. Write press releases about a product | 3.16 | 3.69 | 0.95 | .666 NS |
| 63. Prepare a sample product | 3.56 | 5.06 | 9.34 | .003 |
| 64. Write specifications for user manual | 3.74 | 4.69 | 3.28 | .070 NS |
| 65. Evaluate the product's distribution | 2.77 | 3.75 | 3.77 | .053 NS |
| 66. Evaluate effect of a demonstration | 3.77 | 4.69 | 3.80 | .052 NS |
| 67. Identify parts of a marketing study | 3.37 | 4.06 | 1.55 | .214 NS |
| 68. Interview users about product use | 3.88 | 5.12 | 5.97 | .016 |
| 69. Establish cooperation with users | 4.15 | 5.50 | 7.34 | .008 |
| 70. Deliver an oral presentation | 4.11 | 5.38 | 5.58 | .020 |
| 71. Demonstrate the use of a product | 4.53 | 5.19 | 1.64 | .202 NS |
| 72. Translate data into verbal form | 3.81 | 4.38 | 1.06 | .308 NS |
| TOTAL - DISSEMINATION | 3.44 | 4.47 | 8.41 | .00051 |

* - All Values tabled .000 have p-Levels of .0005 or less

NS - Not significant at 5% level

It is noted that where 20 of the development items and 16 of the field test and evaluation items are rated 5.0 or higher by the developer group, only 8 of the dissemination and marketing items are rated above 5.0.

Analysis of Scales

Originally the DD&E Diagnostic Instrument was designed to provide the three context scores, six process skill scores, and a total score, but the cluster analysis indicated that there were four major clusters. To determine the relationship among the several possible scales, the scale scores were correlated and factor analyzed. The group to which the subject belonged, as described on p. 4 , was included as a variable roughly indexing development experience. Table 3 presents the results including scale means and standard deviations, the average proficiency rating for items in the scale (item mean), the table of intercorrelations, and the results of the factor analysis.

Because the number of items in the scales is different, the item means have been included in the table to facilitate comparison. The item mean for the Total Score is 4.02. Among the three context scales, the dissemination and marketing scale is markedly lower (3.65) than the other two scales (4.28, 4.14). Among the skill scales, communication skills are highest (4.68) followed by collecting information (4.12). This is not surprising since most college students have some proficiency in at least some of the items included in these two scales. The lowest mean is for planning skills (3.66). Among the four clusters, we find results that not unexpectedly reflect those obtained for the context scales. Again, development is highest (4.3), but the low cluster is publication, production, and public relations (3.26).

The standard deviations (S.D.), also reflect the number of items in the scale. We haven't included item S.D.'s, but it can be noted that the S.D.'s for

TABLE 3

DD&E DIAGNOSTIC INSTRUMENT - MEANS, STANDARD DEVIATIONS, ITEM MEANS, AND FACTOR ANALYSIS OF CONTEXT, PROCESS SKILL, AND CLUSTER SCALES

Decimals omitted for correlations and factor leadings

(N=78)

| | Scale Mean | Scale S.D. | Item Mean | Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | Variable | Factor I | Factor II | Factor III | Factor IV | % C. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Development Context | 102.8 | 29.6 | 4.28 | 1 | - | 76 | 79 | 84 | 90 | 90 | 89 | 86 | 86 | 93 | 80 | 84 | 79 | 94 | 45 | 1 | 59 | 50 | 56 | 21 | 96 |
| Field Test and Evaluation Context | 99.2 | 29.2 | 4.14 | 2 | 76 | - | 58 | 85 | 84 | 77 | 71 | 87 | 76 | 65 | 99 | 64 | 61 | 86 | 36 | 2 | 20 | 91 | 32 | 12 | 98 |
| Dissemination and Marketing Context | 87.5 | 31.6 | 3.65 | 3 | 79 | 58 | - | 72 | 80 | 90 | 85 | 79 | 87 | 70 | 56 | 98 | 88 | 89 | 31 | 3 | 28 | 25 | 90 | 11 | 97 |
| Collecting Information Skills | 49.4 | 14.4 | 4.12 | 4 | 84 | 85 | 72 | - | 84 | 81 | 71 | 81 | 81 | 71 | 85 | 77 | 73 | 89 | 40 | 4 | 20 | 72 | 52 | 18 | 86 |
| Analyzing Skills | 46.0 | 14.4 | 3.83 | 5 | 90 | 84 | 80 | 84 | - | 90 | 83 | 89 | 80 | 84 | 85 | 84 | 76 | 94 | 43 | 5 | 45 | 62 | 54 | 20 | 91 |
| Planning Skills | 43.9 | 14.8 | 3.66 | 6 | 90 | 77 | 90 | 81 | 90 | - | 88 | 87 | 87 | 81 | 77 | 93 | 84 | 96 | 39 | 6 | 40 | 48 | 72 | 15 | 93 |
| Producing Skills | 46.5 | 15.4 | 3.87 | 7 | 89 | 71 | 85 | 71 | 83 | 88 | - | 81 | 84 | 83 | 72 | 86 | 84 | 91 | 36 | 7 | 50 | 39 | 68 | 13 | 88 |
| Evaluating Skills | 47.6 | 14.4 | 3.96 | 8 | 86 | 87 | 79 | 81 | 89 | 87 | 81 | - | 83 | 78 | 89 | 83 | 69 | 93 | 39 | 8 | 38 | 69 | 51 | 14 | 89 |
| Communicating Skills | 56.2 | 14.3 | 4.68 | 9 | 86 | 76 | 87 | 81 | 80 | 87 | 84 | 83 | - | 73 | 74 | 89 | 85 | 92 | 34 | 9 | 23 | 50 | 76 | 11 | 89 |
| Development Cluster | 38.7 | 12.6 | 4.30 | 10 | 93 | 65 | 70 | 71 | 84 | 81 | 83 | 78 | 73 | - | 69 | 76 | 62 | 85 | 41 | 10 | 79 | 37 | 41 | 18 | 98 |
| Field Test and Evaluation Cluster | 102.6 | 30.9 | 4.10 | 11 | 80 | 99 | 56 | 85 | 85 | 77 | 72 | 89 | 74 | 69 | - | 63 | 58 | 86 | 38 | 11 | 28 | 90 | 27 | 15 | 99 |
| Dissemination Cluster | 93.5 | 31.5 | 3.89 | 12 | 84 | 64 | 98 | 77 | 84 | 93 | 86 | 83 | 89 | 76 | 63 | - | 83 | 92 | 37 | 12 | 34 | 32 | 84 | 15 | 94 |
| Publication, Production and P.R. Cluster | 43.4 | 16.3 | 3.62 | 13 | 79 | 61 | 88 | 73 | 76 | 84 | 84 | 69 | 85 | 62 | 58 | 83 | - | 85 | 30 | 13 | 15 | 31 | 87 | 10 | 89 |
| TOTAL SCORE | 289.5 | 81.2 | 4.02 | 14 | 94 | 86 | 89 | 89 | 94 | 96 | 91 | 93 | 92 | 85 | 86 | 92 | 85 | - | 41 | 14 | 39 | 61 | 67 | 16 | 100 |
| Group 1=St, 2=G. St, 3=Dev, 4=Sr. Dev. | 2.12 | .69 | 2.12 | 15 | 45 | 36 | 31 | 40 | 43 | 39 | 36 | 39 | 34 | 41 | 38 | 37 | 30 | 41 | - | 15 | 13 | 18 | 14 | 97 | 100 |
| Percent of trace | | | | | | | | | | | | | | | | | | | | | 16 | 31 | 39 | 08 | - |

the three context scales (each 24 items) are quite similar, as are the S.D.'s for the six skill scales (each 12 items).

The table of intercorrelations indicates that all of the scales are highly intercorrelated. Among the three context scales, the lowest correlation is .58 between the dissemination and marketing and the field test and evaluation scales. Both of these scales correlated with the development scale in the high .70's. All of the process skills correlated .71 or higher. Among the cluster scales, the lowest correlation is .58 between the field test and evaluation cluster scale and the publication, production, and public relations cluster scale. The .83 correlation between this latter scale and the dissemination cluster scale indicates that these two cluster scales are substantially correlated (and in fact appeared as a single factor in the factor analysis). The reader should recall that the three sets of scales (context, process skills, and clusters) are all derived from the same set of 72 items and thus should show substantial intercorrelation, but it is perhaps surprising that the lowest between scale-type correlation is .61 between the field test context and the publication, production, and public relations cluster.

None of the scales correlated highly with the group index (1=student, 2=graduate student, 3=developer, and 4=senior developer), although it should be recalled that 51 of the 78 subjects were in one group, graduate students. All of the correlations between scales and the group index are between .30 and .45.

A principal components factor analysis was performed. The first factor extracted 79 percent of the trace, and all loading for the scales were above .85, indicating that a single general proficiency factor could account for most of the covariance among the scales. The second factor, accounting for seven percent of the trace distinguished among the three contexts with loadings of .41 on field test and evaluation, .02 on development, and -.41 on evaluation. The third factor accounting for 5 percent of the trace was associated almost solely with the group index. It had a loading of .79 on the group index. The only scales with even modest loading were variable 2 (-.27) and variable 11 (-.23), both concerned with field test and evaluation. A fourth factor, accounting for only three percent of the trace, was defined in terms of differences between development (variable 10 loading -.48 and variable 1 loading -.21) and publication, production, and public relations (variable 13 loading +.22).

These four factors, accounting for 94 percent of the trace, were rotated by the Varimax method (Kaiser) to obtain the four orthogonal factors reported in Table 3 . Factor I is the development factor, which is best identified by the .79 loading on variable 10 and the .59 loading on variable 1. Factor II is the field test and evaluation factor, identified by the .91 loading on variable 2, and the .90 loading on variable 11. Factor III, a production and dissemination factor, identified by the .90 loading on variable 3, the .84 loading on variable 12 and the .87 loading on variable 13. Factor IV is almost solely associated with the group index variable, which loads on this factor .97. The four rotated factors account for the following percentage of the trace respectively: 16%, 31%, 39%, and 8%. Excluding Factor IV, we see that development (Factor I) is least well represented among the three DD&E factors. This result is paralleled in terms of the loadings of the Total Score, variable 14, on the first three factors (.39, .61, .67). These results strongly indicate that the DD&E Diagnostic Instrument is deficient in statements that reflect competencies of the type contained in the development cluster. On the other hand, both evaluation (Factor II) and dissemination (Factor III) are well represented.

Turning to individual scales, we note that the development context scale loads approximately evenly on all of the first three factors. This result is comparable to the finding discussed on p 7 regarding the cluster analysis. In the current DD&E matrix, the development column is a mixture of several types of competence statements. By contrast, variables 2 and 3 exhibit relatively pure, but not completely uncontaminated factor composition.

Skipping to the four cluster variables, because they are most closely related to the context variables, we note that the factor composition is well defined, but not quite so pure. The development cluster, variable 10, does have a much stronger loading on Factor I (.79) than on the other two factors (.37 and .41). The field test and evaluation cluster (variable 11) has a loading pattern quite similar to variable 2, but this should be expected since they share a large number of items. Reflecting the .83 correlation between variables 12 and 13, we find that these two variables exhibit highly similar patterns of factor loadings; both heavily loaded on Factor III, and with much smaller loadings on Factors I and II.

Turning finally to the six skills scales, we see that the collecting information skills scale, the analyzing skills scale, and the evaluation skills scale show highest loadings on Factor II, evaluation; while the remaining three skill scales, planning, producing, and communication, show highest loadings on Factor III, production and dissemination. Although Factor I, development, is never highest, it does have modest loadings on at least four of the skill scales, with the highest (.50) for producing skills, and the next highest (.45) for analyzing skills. These results indicate that although the process skills scales are highly intercorrelated (.71 to .90), they do exhibit somewhat different, if mixed, factor patterns. The patterns seem to make some "common sense" in that collecting information skills, analyzing skills, and evaluating skills are most prominantly associated with Factor II, the evaluation factor; whereas planning, producing and communicating tend to be associated with Factor III, production and dissemination. None of the six process skills display loadings above .50 on Factor I, development, but at least four process skills--analyzing, planning, producing, and evaluating--have loadings above .38. Table 4 below presents the analysis of differences between the student and the developer groups on each of the scales.

TABLE 4

PROFICIENCY LEVELS FOR STUDENTS AND FOR DEVELOPERS BY SUBSCALES

| SCALE | STUDENTS (N=62) | DEVELOPERS (N=16) | F | p |
|---|---|---|---|---|
| Development Context | 3.99 | 5.39 | 20.12 | .0001 |
| Field Test and Evaluation Context | 3.86 | 5.22 | 19.59 | .0001 |
| Dissemination and Marketing Context | 3.44 | 4.47 | 8.41 | .0051 |
| *** | | | | |
| Collecting Information Skills | 3.84 | 5.18 | 19.37 | .0001 |
| Analyzing Skills | 3.55 | 4.94 | 21.41 | .0001 |
| Planning and Designing Skills | 3.41 | 4.64 | 14.58 | .0005 |
| Producing Skills | 3.62 | 4.85 | 13.38 | .0008 |
| Evaluating Skills | 3.69 | 5.00 | 18.19 | .0002 |
| Communicating Skills | 4.45 | 5.55 | 12.24 | .0011 |
| *** | | | | |
| Development Cluster | 4.04 | 5.33 | 12.35 | .0011 |
| Field Test and Evaluation Cluster | 3.81 | 5.25 | 21.66 | .0001 |
| Dissemination and Marketing Cluster | 3.65 | 4.86 | 12.25 | .0011 |
| Publication, Production, and Public Relations Cluster | 3.40 | 4.49 | 9.00 | .0039 |
| TOTAL SCORE | 3.76 | 5.03 | 19.55 | .0001 |

## INSTRUMENT REVISION

The various data analyses previously described (i.e., cluster analysis and factor analysis of scales), designed to determine if the assumed "context by skills" matrix organization could be supported empirically, found that though the DD&E Diagnostic Instrument items were highly intercorrelated, there were at least three well-defined clusters associated with the three DD&E contexts. Though the analyses failed to produce any clusters associated with a single process skill (the rows in the matrix of Figure 1), the clusters exhibited markedly different patterns of process skills. So, although 1) the process skills did not emerge cleanly, 2) two highly correlated clusters associated with the dissemination context were found, and 3) the items in the development cluster were only a small (n=9) subset of the 24 items in the development context, the data analysis did lend empirical support retention of the structure of the DD&E matrix. Given the preliminary nature of these results and also the fact that our sample of 78 test subjects involved only 16 developers, we decided to keep the structure of our *a priori* matrix of three contexts and six process skills, and to focus our revision on the refinement of items. It should be noted that most of the revisions were based on the cluster analysis results, on the review of appropriateness for specific cells of the DD&E matrix, and on the levels and distribution of proficiency ratings. Only two items were changed on the basis of significance of differences between the student and the developer groups. In both cases, the items were rewritten to describe more specific competencies.[3]

---

[3] Selection of items on the basis of discrimination would be more appropriate for a *norm* referenced instrument. The DD&E instrument is primarily *criterion* referenced Its primary purpose is not to discriminate between groups with different general levels of experience, but to identify the levels of proficiency for specific competency areas.

## Revisions Based on Cluster Analysis

The four clusters were studied to identify items which did not cluster with other items in their column (development, evaluation, or dissemination). These items were then reviewed to determine if they should be revised, moved to a different column, or replaced with a new item. For example, item 13 is a development item which appeared in a dissemination cluster. The original item was:

> Take steps to assure that such things as waivers, releases, copyright releases, or patent protections are secured when appropriate.

After reviewing the item itself, and other items in the cluster, it was determined that item 13 was (1) describing an activity often considered a dissemination activity, and (2) very similar to item 56 which was already in the dissemination column. The original item 13 was replaced with a development activity:

> Prepare scripts for instructional films or sound filmstrips.

Another example would be item 45, an evaluation item which was grouped into a dissemination cluster. In its original form the item was worded:

> Prepare an article for publication in a scholarly professional journal.

In this form, the content of the article has not been specified. The item was reworded so that it would represent an evaluation activity:

> Prepare an article describing an evaluation study for publication in a professional journal.

## Revisions Based on Item Location

All items were reviewed to determine if the activity they described was appropriate for the DD&E Matrix cell in which they had been placed. Items which did not clearly describe an activity of their cell were revised or replaced with a new activity statement. For example, the original activity described in item 43 was:

> Determine the internal consistency reliability of a knowlege test instrument.

Because this item appears in the evaluation process skill row, it should describe an activity requiring a critique or review which includes a judgment of value. Since the original item does not require judgment or decision, it was replaced. The new item 43 is:

> Critique a field test plan in terms of technical adequacy, feasibility, and cost effectiveness.

A list of items in the revised DD&E Diagnostic Instrument can be found on the following pages. New or revised items are indicated by an asterisk (*).

## REVISED DIAGNOSTIC INSTRUMENT

### Development of Educational Products

*Collecting Information*

*1. Use bibliographic resources such as Research in Education and Current Index to Journals in Education to locate information supporting the need to develop an educational program or product.

2. Locate existing methods or strategies which can be used in potential product development.

*3. Conduct a search to determine if products exist which could meet your need(s) for instructional materials.

*4. In preparing a proposal for a development project, search for and select references which would support the rational and technical approach of the project.

*Analyzing*

5. Given a problem statement, information on the history of the problem, objectives and possible solutions, write a 20-page proposal for solving the problem, including a rationale for the approach and a development schedule.

*6. Use theories of instruction and learning to design an educational product or program.

*7. Analyze job or task requirements to determine objectives for an instructional program or product.

8. Categorize instructional objectives in terms of a taxonomy (e.g., Bloom's Taxonomy of the cognitive domain or Gagne's conditions of learning).

*Planning*

*9. Plan the budget and schedule for a development project.

10. In producing specifications for instructional materials determine an appropriate format for the materials.

11. Arrange learning activities in a sequence to facilitate learning or mastery of objectives.

*12. Design alternative instructional methods for attaining the same set of objectives.

---

*New or revised item.

*Producing*

*13. Prepare scripts for instructional films or sound filmstrips.

*14. Improve curricular materials by revising them according to instructional technology principles.

15. Write exercises which the learner should do in order to master concepts or principles in an instructional unit.

16. Write copy for instructional materials from product specifications.

*Evaluating*

*17. Review first draft materials to determine if they comply with specifications regarding objectives, sequencing, and content.

*18. Evaluate alternative methods of presenting instructional material/media and instructional methods to match them with instructional objectives and learner characteristics.

*19. Provided with product specifications, the test product, field test reports, and expert reviews, make recommendations regarding the nature and extent of revisions required.

*20. Evaluate the feasibility and risks associated with alternative approaches for redesigning a product that has failed to meet specifications.

*Communicating*

*21. Prepare a technical report stating the need, rationale, and proposed technical approach for developing an educational product or program.

*22. Prepare a memorandum which summarizes the actions taken and problems encountered in a meeting with your development team.

*23. Secure the services of, and meet with consultants to obtain their advice and recommendations regarding a development project.

*24. Make a formal, oral report to the administrators of your organization regarding the status of a development project.

## Field Test and Evaluation

### *Collecting Information*

25. Prepare a coding scheme which will be used by a group of coders in preparing field test data for computer analysis.

26. Organize statistical data into a meaningful presentation.

27. Interpret scatter plots.

28. Design data collection procedures to maintain privacy or confidentiality in collecting, processing and storing information.

### *Analyzing*

29. List the major factors which jeopardize the internal and external validity of a specified evaluation study.

30. Formulate significant, answerable questions for an evaluation study.

31. Analyze discrepancies between expected and actual test outcomes.

32. Determine if new test instruments need to be developed by reviewing how well available tests fit the evaluation objectives.

### *Planning*

33. Determine which standard procedure for establishing validity is best for your test instrument.

34. Design a monitoring system that will provide data on the status of the operating system (such as actual vs. intended outcomes, unmet needs, problems, etc.).

35. Given a situation where a randomly assigned control group can not be established, suggest feasible methods to control for extraneous variables that may confuse the evaluation results.

*36. In planning the evaluation of a new training program, determine the types of subjects which should be used.

*Producing*

*37. Revise a test instrument on the basis of data collected during a pilot test of the instrument.

38. Make adjustments to test administration procedures, when situational factors make such adjustments essential, in a manner that will secure and protect the validity of the most important data.

*39. Specify a set of procedures to be followed by evaluation personnel in administering test instrument.

40. Explain the importance of standardized procedures in conducting an interview.

*Evaluating*

41. Determine if theoretical assumptions underlying various statistical techniques have been violated in analysis of data.

*42. Evaluate proposed test instruments to determine if they are valid, reliable and appropriate for specific evaluation purposes.

*43. Critique a field test plan in terms of technical adequacy, feasibility and cost effectiveness.

*44. Review a draft field test report to determine if it is correct, complete and ready for release.

*Communicating*

*45. Prepare an article describing an evaluation study for publication in a professional journal.

*46. Give an oral report of the findings of an evaluation study.

47. Prepare simple evaluation reports summarizing findings and interpretations of collected data.

*48. Prepare graphs to display numerical information summarizing field test results.

## Dissemination and Marketing

### *Collecting Information*

49. Collect information on installation costs for your instructional product.

50. Design data collection forms for a marketing study of an educational product.

*51. Obtain information about strategies used to market educational products similar to yours.

*52. Retrieve information on political, social, and economic factors which may have a bearing on the dissemination and marketing of an educational product.

### *Analyzing*

53. Identify the crucial characteristics of a target group which may influence the dissemination effort.

54. Evaluate at least three different market research techniques applicable to a specific development.

55. Review alternatives for the design of a product in terms of possible problems in installing or maintaining the product.

56. Make arrangements to secure copyrights and copyright clearances where needed.

### *Planning*

57. Outline factors which must be considered in disseminating information about an educational product designed for a specific target audience.

58. Design an effective public relations activity for a research and development organization or project.

59. Plan interviews with potential users for the purpose of determining a market for your product.

60. Prepare specifications for audio-visual materials which will be used in the dissemination effort.

### *Producing*

61. Write public relations scripts for film and slide shows.

62. Write press releases to disseminate information about a new product.

63. Prepare a sample of an educational product which can be presented to a specified target group in order to determine if the proposed product meets the group's needs.

*64. Prepare a brochure for users which describes maintenance and proper usage of an installed product.

### *Evaluating*

65. Determine the thoroughness of distribution which occurred in dissemination of an educational product.

*66. Determine the effectiveness of using a demonstration to disseminate information to your target audience.

*67. Given the results of a marketing study, review alternative dissemination plans to determine which is most appropriate for the product and market studied.

68. After product installation, interview users to determine if adequate information was provided about how to install and use the product.

### *Communicating*

*69. Interact with users to help them install a new educational product or program.

70. Deliver an oral presentation to a group of more than 20 professional educators to disseminate information about a project or product.

*71. Confer with different types of users about their problems in the use of and educational product.

*72. Translate technical information from a marketing study into a readable summary report.

# APPENDIX A

# REVISED EIGHT-POINT RATING SCALE AND ORIGINAL DIAGNOSTIC INSTRUMENT

## APPENDIX A

## REVISED EIGHT-POINT RATING SCALE

Note: Interviews with subjects and analysis of frequency counts of levels of proficiency reported indicate that some subjects, especially the highly experienced professionals, needed an eighth level. The revised instrument to be used for discriminant validation has been modified to include this additional statement.

### Level of Proficiency

1. I have no specific knowledge about this activity nor experience with it.
2. I have read about or seen this activity performed, but have no experience with it and don't really understand it.
3. I have studied this activity or have frequently seen it performed and have a good understanding of it, *but I have not yet done it.*
4. I have a general understanding of this activity and have had some experience with it, enough so that I can do it if I have either detailed instructions or close supervision.
5. I have enough experience in performing this activity to do it if given enough general supervision or general instructions.
6. I have enough knowledge and experience with this activity so that I can perform this task quite satisfactorily without supervision or job aids.
7. I have had extensive experience with this activity, and can perform it quickly, efficiently, and do a top quality job.
8. I consider myself an expert in this activity and fully qualified to accomplish unusually difficult or completely novel work.

## ORIGINAL DIAGNOSTIC INSTRUMENT

### Development of Educational Products

1. Retrieve information on political, social or economic factors which have a bearing on the analysis of an educational problem.

2. Locate existing methods or strategies which can be used in potential product development.

3. Prior to conducting a survey of the literature, prepare a search strategy.

4. In the context of conducting a survey of literature, scan and evaluate obtained information for relevance.

5. Given a problem statement, information on the history of the problem, objectives and possible solutions, write a 20-page proposal for solving the problem, including a rationale for the approach and a development schedule.

6. Provided with current theories of instruction, relate them to the formulation of a design for an educational product or program.

7. Provided with field test data on instructional materials, examine low gain scores and determine if they indicate problems in test construction or instructional materials.

8. Categorize instructional objectives in terms of a taxonomy (e.g., Bloom's Taxonomy of the cognitive domain or Gagne's conditions of learning).

9. Given funding resources, time and cost estimates, and project priorities plan a budget.

10. In producing specifications for instructional materials determine an appropriate format for the materials.

11. Arrange learning activities in a sequence to facilitate learning or mastery of objectives.

12. Confer with production personnel to determine materials needed.

13. Take steps to assure that such things as waivers, releases, copyright releases or patent protections are secured when appropriate.

14. Confer with specialists when problems of production cannot be solved on the job.

15. Write exercises which the learner should do in order to master concepts or principles in an instructional unit.

16. Write copy for instructional materials from product specifications.

17. Translate field test data into recommendations for action.

18. Conduct a case study of a program or project.

19. Provided with product specifications, review a product and documentation on product development and field testing to determine if the specifications have been met.

20. Informally try out a development product with one or only a small group of subjects to observe and record how the test subjects use the materials, where and why they have difficulties, etc.

21. Write a position paper justifying the development of an educational product.

22. Discuss the development of an educational product with potential user-groups.

23. Interact and contribute in a staff or consultant meeting.

24. Communicate product specifications to personnel who will be producing the product.

## Field Test and Evaluation

25. Prepare a coding scheme which will be used by a group of coders in preparing field test data for computer analysis.

26. Organize statistical data into a meaningful presentation.

27. Interpret scatter plots.

28. Design data collection procedures to maintain privacy or confidentiality in collecting, processing and storing information.

29. List the major factors which jeopardize the internal and external validity of a specified evaluation study.

30. Formulate significant, answerable questions for an evaluation study.

31. Analyze discrepancies between expected and actual test outcomes.

32. Determine if new test instruments need to be developed by reviewing how well available tests fit the evaluation objectives.

33. Determine which standard procedure for establishing validity is best for your test instrument.

34. Design a monitoring system that will provide data on the status of the operating system (such as actual vs. intended outcomes, unmet needs, problems, etc.).

35. Given a situation where a randomly assigned control group can not be established, suggest feasible methods to control for extraneous variables that may confuse the evaluation results.

36. Determine criteria for selection of field test sites for a specified educational product.

37. Make revisions in test instruments based on evaluation data.

38. Make adjustments to test administration procedures, when situational factors make such adjustments essential, in a manner that will secure and protect the validity of the most important data.

39. Prepare a test administration manual.

40. Explain the importance of standardized procedures in conducting an interview.

41. Determine if theoretical assumptions underlying various statistical techniques have been violated in analysis of data.

42. Evaluate test instruments using data collected in try-out and revision cycles in order to recommend instrument revisions for the final field test.

43. Determine the internal consistency reliability of a knowledge test instrument.

44. Given a report on the evaluation of an educational product, identify the purpose of the evaluation and the steps used in the process.

45. Prepare an article for publication in a scholarly professional journal.

46. Give a short speech or oral report.

47. Prepare simple evaluation reports summarizing findings and interpretations of field test data.

48. Prepare graphs to display numerical information.

## Dissemination and Marketing

49. Collecting information on installation costs for your instructional product.

50. Design data collection forms for a marketing study of an educational product.

51. Use at least three different information services to obtain information about educational products similar to yours.

52. Construct an annotated bibliography to accompany a brochure describing the product being disseminated.

53. Identify the crucial characteristics of a target group which may influence the dissemination effort.

54. Evaluate at least three different market research techniques applicable to a specific development.

55. Review alternatives for the design of a product in terms of possible problems in installing or maintaining the product.

56. Make arrangements to secure copyrights and copyright clearances where needed.

57. Outline factors which must be considered in disseminating information about an educational product designed for a specific target group.

58. Design an effective public relations activity for a research and development organization or project.

59. Plan interviews with potential users for the purpose of determining a market for your product.

60. Prepare specifications for audio-visual materials which will be used in the dissemination effort.

61. Write public relations scripts for film and slide shows.

62. Write press releases to disseminate information about a new product.

63. Prepare a sample of an educational product which can be presented to a specified target group in order to determine if the proposed product meets the group's needs.

64. Prepare the specifications for a manual which describes maintenance and proper usage of an installed product.

65. Determine the thoroughness of distribution which occurred in dissemination of an educational product.

66. Carry out an evaluation of the effectiveness of a demonstration of an educational product.

67. Given a marketing study identify: a) the problem, b) how the study was designed and c) the outcomes of the study.

68. After product installation, interview users to determine if adequate information was provided about how to install and use the product.

69. Interact with users to establish better understanding and cooperation in installing a new educational product or program.

70. Deliver an oral presentation to a group of more than 20 professional educators to disseminate information about a project or product.

71. Conduct demonstrations on the use of an educational product.

72. Translate quantitative or numerical information from a marketing study into verbal or narrative form.

# APPENDIX B

# RESULTS OF THE CLUSTER ANALYSIS

# APPENDIX B

Results of the cluster analysis are presented below. For each cluster, the items are listed in the order in which they were added to the cluster. The letter to the left of the original item number indicates the DD&E matrix colum in which the item was located. D = Development, E = Evaluation and M = Dissemination/Marketing.

## Cluster I - Publication, Production and Public Relations

M 60. Prepare specifications for audio-visual materials which will be used in the dissemination effort.

M 61. Write public relations scripts for film and slide shows.

M 62. Write press releases to disseminate information about a new product.

M 56. Make arrangements to secure copyrights and copyright clearances where needed.

D 13. Take steps to assure that such things as waivers, releases, copyright releases or patent protections are secured when appropriate.

M 58. Design an effective public relations activity for a research and development organization or project.

M 70. Deliver an oral presentation to a group of more than 20 professional educators to disseminate information about a project or product.

D 21. Write a position paper justifying the development of an educational product.

D 12. Confer with production personnel to determine materials needed.

Cluster I (continued)

E 45. Prepare an article for publication in a scholarly professional journal.

D 9. Given funding resources, time and cost estimates, and project priorities plan a budget.

D 1. Retrieve information on political, social or economic factors which have a bearing on the analysis of an educational problem.

## Cluster II - Field Test and Evaluation

E 42. Evaluate test instruments using data collected in try-out and revision cycles in order to recommend instrument revisions for the final field test.

E 37. Make revisions in test instruments based on evaluation data.

D 17. Translate field test data into recommendations for action.

E 39. Prepare a test administration manual.

D 7. Provided with field test data on instructional materials, examine low gain scores and determine if they indicate problems in test construction or instructional materials.

E 32. Determine if new test instruments need to be developed by reviewing how well available tests fit the evaluation objectives.

D 20. Informally try out a development product with one or only a small group of subjects to observe and record how the test subjects use the materials, where and why they have difficulties, etc.

E 43. Determine the internal consistency reliability of a knowledge test instrument.

D 6. Provided with current theories of instruction, relate them to the formulation of a design for an educational product or program.

E 31. Analyze discrepancies between expected and actual test outcomes.

E 38. Make adjustments to test administration procedures, when situational factors make such adjustments essential, in a manner that will secure and protect the validity of the most important data.

Cluster II (continued)

E 25. Prepare a coding scheme which will be used by a group of coders in preparing field test data for computer analysis.

E 33. Determine which standard procedure for establishing validity is best for your test instrument.

E 47. Prepare simple evaluation reports summarizing findings and interpretations of field test data.

E 34. Design a monitoring system that will provide data on the status of the operating system (such as actual vs. intended outcomes, unmet needs, problems, etc.).

E 30. Formulate significant, answerable questions for an evaluation study.

E 35. Given a situation where a randomly assigned control group can not be established, suggest feasible methods to control for extraneous variables that may confuse the evaluation results.

D 18. Conduct a case study of a program or project.

E 28. Design data collection procedures to maintain privacy or confidentiality in collecting, processing, and storing information.

E 26. Organize statistical data into a meaningful presentation.

E 41. Determine if theoretical assumptions underlying various statistical techniques have been violated in analysis of data.

E 27. Interpret scatter plots.

E 40. Explain the importance of standardized procedures in conducting an interview.

D 4. In the context of conducting a survey of literature, scan and evaluate obtained information for relevance.

E 29. List the major factors which jeopardize the internal and external validity of a specified evaluation study.

## Cluster III - Dissemination (Planning and Evaluation)

M 57. Outline factors which must be considered in disseminating information about an educational product designed for a specific target group.

M 66. Carry out an evaluation of the effectiveness of a demonstration of an educational product.

M 59. Plan interviews with potential users for the purpose of determining a market for your product.

M 55. Review alternatives for the design of a product in terms of possible problems in installing or maintaining the product.

M 65. Determine the thoroughness of distribution which occurred in dissemination of an educational product.

M 68. After product installation, interview users to determine if adequate information was provided about how to install and use the product.

E 36. Determine criteria for selection of field test sites for a specified educational product.

M 63. Prepare a sample of an educational product which can be presented to a specified target group in order to determine if the proposed product meets the group's needs.

M 54. Evaluate at least three different market research techniques applicable to a specific development.

M 71. Conduct demonstrations on the use of an educational product.

M 51. Use at least three different information services to obtain information about educational products similar to yours.

Cluster III (continued)

D 22. Discuss the development of an educational product with potential user-groups.

M 64. Prepare the specifications for a manual which describes maintenance and proper usage of an installed product.

D 24. Communicate product specifications to personnel who will be producing the product.

D 14. Confer with specialists when problems of production cannot be solved on the job.

M 69. Interact with users to establish better understanding and cooperation in installing a new educational product or program.

M 67. Given a marketing study identify: a) the problem, b) how the study was designed and c) the outcomes of the study.

M 49. Collect information on installation costs for your instructional product.

M 53. Identify the crucial characteristics of a target group which may influence the dissemination effort.

E 44. Given a report on the evaluation of an educational product, identify the purpose of the evaluation and the steps used in the process.

M 50. Design data collection forms for a marketing study of an educational product.

M 52. Construct an annotated bibliography to accompany a brochure describing the product being disseminated.

M 72. Translate quantitative or numerical information from a marketing study into verbal or narrative form.

D 23. Interact and contribute in a staff or consultant meeting.

## Cluster IV - Development

D 11. Arrange learning activities in a sequence to facilitate learning or mastery of objectives.

D 15. Write exercises which the learner should do in order to master concepts or principles in an instructional unit.

D 19. Provided with product specifications, review a product and documentation on product development and field testing to determine if the specifications have been met.

D 8. Categorize instructional objectives in terms of a taxonomy (e.g., Bloom's Taxonomy of the cognitive domain or Gagne's conditions of learning).

D 10. In producing specifications for instructional materials determine an appropriate format for the materials.

D 16. Write copy for instructional materials from product specifications.

D 2. Locate existing methods or strategies which can be used in potential product development.

D 5. Given a problem statement, information on the history of the problem, objectives and possible solutions, write a 20-page proposal for solving the problem, including a rationale for the approach and a development schedule.

D 3. Prior to conducting a survey of the literature, prepare a search strategy.

## Items Which Did Not Appear in a Cluster

E 48. Prepare graphs to display numerical information.

E 46. Give a short speech or oral report.

# Appendix B

## Development and Validation of DD&E (Development, Dissemination, and Evaluation) Competence Assessment Battery

Laird Blackwell
Paul D. Hood

## A. INTRODUCTION

The Far West DD&E Functional Competence Training Program is a competence-based professional development program which provides training resources (23 instructional modules) aimed at "entry-level" skills and knowledge required in educational development, dissemination, and evaluation (DD&E). One part of the program's effort has focused on the development and validation of an assessment system which is designed to aid both student and instructor in : (1) program planning, (2) progress monitoring, and (3) exit assessment. Although this assessment battery was designed for use with the training program, we see it as having a more general utility as a model assessment system for competence-based professional training programs and as a flexible and adaptable set of assessment instruments that may be employed in any competence-based curriculum whose content includes educational product development and evaluation.

Since the assessment battery is intended to provide the basis for a sequence of reliable decisions (by students and instructors) about classification rather than a one-shot decision about selection, it was felt to be important to develop a variety of instruments that would assess similar skills and knowledge by different methods so that decisions could be based on several sources of information used simultaneously or sequentially. Although this emphasis on a multi-method assessment system has been maintained throughout development; the form, scope, and content of the individual assessment instruments have undergone some changes through the test, review, and revision cycles of formative evaluation.

## B. PROTOTYPE INSTRUMENTS

### 1. Initial Development and Pilot Testing of Assessment Instruments

After review of a number of alternatives, three basic assessment devices were selected for inclusion in the Competence Assessment Battery: self-ratings of competence, knowledge test, and job sample tests. Due to time and money constraints, it was decided to pilot test the proposed assessment system by developing and testing two mini-batteries, each based on the content of one of the

DD&E instructional modules. These mini-batteries could, then, be taken as a model of how assessment of a broader range of DD&E competence could be achieved.

Self-rating items, knowledge test items, and job sample exercises were derived directly from the objectives and instructional content of the two selected DD&E modules. Pilot testing with employees in educational development positions (N=25 and 29 for the two modules respectively) with widely varying levels of experience demonstrated the feasibility of the multi-method battery approach, but left many questions regarding the achievement of convergent and discriminant concurrent validity. Although there were often high intercorrelations between instruments, the restricted content of competencies in the mini-batteries produced a homogeneity which severely limited the possibility of demonstrating discriminant validity. It was decided that future development should cover a much broader range of competencies and that greater care should be exercised in matching the specific content of competencies across instruments.

In addition to the self-rating items specific to the two instructional modules, a "DD&E Diagnostic Instrument" was developed consisting of a broad range of self-rating items derived directly from competencies identified (by survey of professionals as described in Hood, Havassy, Lash, and Ward, 1973) as important for entry-level professionals in any of the areas of educational DD&E. While it was not feasible to develop knowledge tests or job sample tests covering such a broad range of DD&E skills and activities, it was possible to develop a broad self-rating instrument, which could be helpful in student program planning, due to the relative ease and low cost of development, administration, and scoring of this type of assessment instrument. The items on this Diagnostic Instrument were organized in a matrix so as to yield information about proficiency in three contexts (i.e., DD&E) and in six process skills (i.e., collecting information, analyzing, planning, producing and implementing, evaluating, and communicating). After extensive refinement of items based on criticisms by local DD&E professionals, the Diagnostic Instrument was administered to 78 students and professionals in educational DD&E. Analysis of variance indicated significant differences in ratings between students and professionals in the expected direction for almost all the items. Although ratings were highly intercorrelated, the "context by skills" matrix organization was given some support; both cluster analysis and factor analysis produced three clusters of items clearly associated with the three contexts (DD&E) but ratings did not cluster by any one of the six process skills. However, factor analysis revealed different, meaningful patterns of skills loadings on the three DD&E factors. On the basis of this data, some items were revised or refined in preparation for the following field test. Although separate skills had not been validated, the matrix was retained as a device for maintaining reasonable uniformity among the three sets of "DD&E context" items in terms of their representation of all six process skills.

## 2. Development of a Product Development Competence Assessment Battery

In order to focus further effort it was decided to construct and validate just one major assessment battery. The area of "product development" was chosen. The five DD&E Training Program instructional modules in this area cover product design, product engineering, review, tryout, and revision. Although the modules emphasize orientation and familiarization and general procedures employed in product development, and the assessment battery was constructed to reflect the general content of the modules; it was decided not to constrain the content of the battery to conform specifically to the content of the modules. Though focused primarily on development and on entry-level (orientation and familiarization) competencies, the assessment battery was constructed to be considerably broader than the instructional modules both in content and in the level of proficiency assessed.

In addition to the DD& E Diagnostic Instrument which provided self-rating items sampling wide areas of DD&E, a self-rating instrument, a knowledge test, and a job sample test, each addressing the area of "product development," were constructed. The self-rating items were derived from the objectives of the modules as well as from competencies related to product development which experts agreed were appropriate to entry-level DD&E practitioners. These items were similar to those classified as in the "development" context of the DD&E Diagnostic Instrument, but were oriented toward somewhat more specific competencies and activities. The knowledge test items and job samples were derived from: content analysis of the "development" modules, analysis of other R&D instructional and evaluation materials, and expert opinion as to the activities professional developers would likely encounter. Since the job sample exercises were intended to assess "development" skills rather than knowledge about specific subject matter content in which such skills might be exercised, a subject matter area (the Toy Lending Library Project at Far West Laboratory) was chosen which was as simple and straightforward as possible and for which the all necessary background information could be acquired easily and quickly. The "Librarian's Manual" was chosen as the product around which to structure the development activities in the job sample. This manual was to contain instructions for the "librarian" (an adult community member chosen to conduct the training session), teaching parents how to use the library's toys with their children. After considerable review and revision, five job sample exercises were developed: product design, preparation for review, tryout, revision, and outlining a script for an instructional film. Although the "script" exercise did not represent as general an activity as did the other four and might not typically be encountered by entry-level developers, it was included as an example of how job sample exercises might be constructed to assess more specific or technical skills.

The various instruments of the Assessment Battery were intended to be redundant; the self-rating items, knowledge questions, and job sample exercises were designed to assess many of the same competencies by different methods. However, the easier and less expensive the instrument to develop, administer, and evaluate, the broader its scope. Thus, the knowledge test covers some knowledge not relevant to the job sample activities, and the self-rating items address many competencies not covered by either of the other two instruments.

### 3. The Formative Field Test of the Competence Assessment Battery

The purpose of this field test was primarily formative, that is, to gain information that would prove useful in revising the battery. Because the costs of securing suitable test subjects, making arrangements for testing, scoring, and analyzing data are relatively high for a battery of this kind, the sample was somewhat small but was carefully selected to provide a very broad range of relevant experience and training. Thirty-one master's level students from the Educational Development Programs of five universities were recruited with the help of site coordinators as were 19 professionals in Educational DD&E from six R&D agencies. All participants were paid. In addition to this formal testing, a group of expert reviewers were asked to examine the content of the battery. Samples of completed test materials, the scoring system, and other pertinent information were supplied to reviewers in order to help them make substantive and methodological critiques and recommendations.

## 4. Discrimination Between Students and Professionals

### Self-ratings.

Factor analysis of the self-ratings on both the Diagnostic Instrument and the more specific product development self-rating instrument revealed 8 significant factors. Two each related to evaluation and dissemination and four related to development. Although the ratings did cluster according to context (i.e. DD&E),they did not tend to differentiate according to the other dimension ("Process Skills") on the Diagnostic Instrument matrix. when testees were classified into "entry-level students" (N=9), "experienced professionals" (N=6), and "intermediates" (N=25) on the basis of biographic information (e.g., past and present employment, years of DD&E experience, products and publications); self-ratings of "professionals" were higher than those of "students" on all 8 factors. Analyses of variance indicated that these differences were significant ($P < .05$) for all but one of these factors.

### Knowledge test.

The knowledge test was scored according to: whether all requisite information was included ("fraction" score), and how much breadth and imagination was demonstrated ("scope" score). Although the latter score was somewhat subjective, scoring criteria were developed and 85-90% agreement was attained between scores. (The developer of the scoring system did all the scoring of field test responses.) In addition to these three scores described above for the total knowledge test, the same three scores were obtained for several subscales (e.g., tryout, review, script, design) which were derived from factor analysis and examination of the content areas addressed by the knowledge questions. "Professionals" obtained higher scores than did "students" on all subscales and totals. Analyses of variance indicated that these differences were significant ($P < .01$) for the "pass/fail," "fraction," and "scope" scores for the total test, but were significant for only a few of the subscales (e.g., the specialized area of script production.)

### Simulation tasks.

"Pass/fail," "fraction," and "scope" scores were also obtained for each of the five job sample exercises as well as for the total job sample (simulation) test. In addition, several subscales (e.g., consistency of response with the information provided, writing skill, following directions) were derived from factor analysis and examination of the competencies addressed by the exercises. Again, "professionals" obtained higher scores than did "students" on all subscales and totals. Analyses of variance indicated that these differences were significant (P .01 in most cases) for the three total scores, for the scores on three of the individual exercises (review, revision, and script), and for the "consistency with information provided" and "following directions" subscales.

### Multiple correlation regression analysis.

With the two extreme groups of the experience-level classification (i.e., "student" and "professional") as the criterion; the total and subscale scores on the self-ratings, knowledge test, and job sample test accounted for 78% of the variance (i.e., multiple R=.88). The multiple correlation coefficient decreased only slightly when either the knowledge test or job sample test was removed as a predictor (R=.87 and .85 respectively), and decreased somewhat more when only the self-rating was removed (R=.80). Total and subscale scores on each of the three instruments correlated highly with "student vs professional" classification groups (multiple R's ranged from .65 to .77). The patterns of multiple correlations were the same when all three classification groups ("student," "professionals," and "intermediates") were considered, though the correlations were lower. The self-ratings, knowledge test, and job sample test,taken together then,accounted for 45% of the variance in classification (multiple R=.67). This correlation decreased to .53 when only the self-ratings were removed. The patterns of multiple correlations described above suggest that each of the instruments provides powerful discrimination, but that the knowledge test and job sample test may provide somewhat redundant information, while self-ratings provide more independent information.*

---

*Individual comparisons of combinations of predictors in a stepwise multiple regression support this hypothesis, for the (squared) multiple correlation of "knowledge test plus self-ratings" or "job sample test plus self-ratings," with experience classification is significantly different ($P < .05$) than that of "self-rating" alone with experience classification, while the differences between the (squared) multiple correlations of "knowledge test plus job sample test" with experience classification and either test alone with experience classification are not significant.

## 5. Convergent and Discriminant Concurrent Validity

By using an assessment system in which different instruments assess some of the same competencies by different methods, it should be possible to increase the validity of classification decisions which must be made relevant to proficiency on those competencies. In order for this multi-trait, multi-method assessment to be effective, however, convergent and discriminant validity of the system should be demonstrated. Convergent validity would be demonstrated by significant correlations between different methods of measuring the same variable. Although very few significant correlations were found between individual items on the various instruments which presumably address the same competencies; when items were grouped into subscales, almost all the correlations between scores on comparable subscales (i.e., those relating to the same competencies or content areas) for the knowledge test and job sample test were significant ($P < .01$). The three total scores ("pass/fail," "fraction," and "scope") were highly correlated across instruments (the knowledge test/job sample test correlations for these scores were .50, .55, and .52 respectively). There were few significant correlations between self-rating subscales and those of the other two instruments.

According to Campbell and Fiske (1959) there are two ways of demonstrating discriminant validity: (1) higher correlations between measures of the same competence obtained by different methods than between measures of different competencies obtained by different methods, and (2) higher correlations between measures of the same competence obtained by different methods than between measures of different competencies obtained by the same method. There was little of the first type of evidence for discriminant validity — only in some cases were the "same competence, different method" correlations higher than the "different competence, different method" correlations. There was no evidence of the second type to demonstrate discriminant validity — the correlations among subscales on the same instrument were usually high (R=.30 to .80) and were consistently higher than those between comparable subscales on different instruments.

So, although there was evidence that the assessment instruments do discriminate (when used independently or together) between students and professionals and that at least the knowledge test and job sample test are measuring some of the same competencies, there was little evidence of discriminant validity. This might suggest that a strong factor of general ability or intelligence is accounting for some of the variability in scores.

## 6. Review and Revision

The expert reviewers had few comments or criticisms about the form or content of the self-rating items on either the Diagnostic Instrument or the Assessment Instrument more specific to product development. They did, however, indicate that the 8-point "proficiency scale" which testees used to rate their proficiency on the competencies specified in the items was confusing in that it combined levels of knowledge and experience (e.g., "I have read about or seen this activity performed, but have no experience with it and really don't understand it.") Though this combination of knowledge and experience was intentional, we decided to change the scale by developing a 6-point "level of performance" scale which allows testees to rate their performance capabilities primarily in terms of varying amounts of supervision (e.g., "I can perform this activity if I am given enough general supervision or general instruction.").

Since factor analysis and cluster analysis supported the "context" distinctions (i.e., development, dissemination, and evaluation) in the Diagnostic Instrument matrix, but not the "process skills" categories, it didn't seem whorthwhile to retain the matrix organization. Instead, we selected those items on both self-rating forms which had strong factor loadings and combined them into one 60-item self-rating form which still focuses primarily on "product development" (including evaluation) competencies, but which also includes a somewhat broader range of DD&E competencies including those related to dissemination.

Reviewers had few criticisms of the knowledge test. However, on the basis of their comments and on the results of the factor analysis and analysis of variance by student vs professional, several of the questions were refined and a few were eliminated. We deleted all the questions which required definitions of DD&E terms and retained just the short answer questions involving knowledge or application of knowledge in product development and evaluation which we felt was requisite to performance of the job samples. Suggestions for scoring responses were incorporated into the scoring manual which was tried out with local professionals as scorers.

Most reviewers and testees were favorably impressed with the job sample test. Approximately 85% of testees thought the exercises required use of relevent educational product development competencies, and approximately 75% thought they were useful in indicating areas of strength and weakness. However, less than half the testees thought their performance was representative of the quality of work they could do in similar real-life development tasks. This problem was most frequently attributed to shortage of time to work on the exercises. On the basis of these comments and the results of factor analysis and analysis of variance, we eliminated one exercise ("preparation for review") and shortened and refined the others. Since some students complained of too little guidance and some professionals complained of too much, we tried to make the instructions as clear and inclusive as possible without overly constraining possible solutions. According to the reviewers and testees, we were reasonably successful in our attempt to avoid "Far West Laboratory jargon" in the exercises. However, despite efforts to develop a scoring system which would be universally acceptable and useful, our scoring criteria and methods are bound to reflect a particular concept of educational development. For this reason and because of the necessarily high subjective scoring required for complex job samples, we have (in our scoring manual) attempted only to suggest criteria and methods for scoring the attainment of competencies which would allow a reasonably good agreement among scorers. We have emphasized that supervisors or employers could adopt different criteria and/or methods to suit their own purposes and needs.

## C. VALIDATION OF THE REVISED BATTERY

### 1. The Validation Field Test

Sixty-six subjects were recruited from ten different sites, including 26 subjects from four univeristy locations, and 40 subjects from six educational R&D agencies. The R&D agencies included two educational laboratories, one university-based R&D agency, two non-profit R&D organizations, and one government R&D agency.*

---

*To avoid a possible bias, none of the validation subjects were Far West Laboratory employees.

Since the purpose of the validation study was to explore the capability of the battery to discriminate over a wide range of R&D experience, we asked contact persons at each of ten sites to help us in recruiting test subjects with a wide range of experience in educational product development. All participants were paid an honorarium.

In terms of degree of experience the validation subjects may be divided into three groups: 13 subjects with no education R&D work experience and little or no record of authoring any kind of educational R&D publication or product; 21 subjects with modest R&D work experience and some record of producing publications and products; and 32 subjects with significant R&D work experience and a substantial record of publication and product development. Since the Far West DD&E Functional Competence Program is focused at entry-level competence, which we define as M.A. degree plus one to two years of experinece, the second ("modest experience") group is closest to the program's target population. The other two groups provide a "bracket" which help to define a broad range of DD&E competence.

As indicated in Table 1, there are major differences among the groups in age, in years of R&D work experience, in number of publications and products produced, and in level of education.* Additionally, the high R&D experience group is predominantly male; substantially more of the modest R&D experience group have some teaching experience.

As we see in the next section, these three groups also differ markedly in terms of their self-appraisal of R&D competencies.

---

*The modest R&D experience group includes three doctoral level subjects. Two were in academic departments; the third was a recent employee of an R&D agency. Because of their relative lack of significant R&D work experience and very modest record of publications and products, they resembled the modest R&D experience group more than the high experience group.

## TABLE B1

## BIOGRAPHIC DATA ON THE VALIDATION GROUPS

| | R&D EXPERIENCE AND PRODUCTIVITY | | |
|---|---|---|---|
| | None | Modest | High |
| Number in group | 13 | 21 | 32 |
| Age (years) | 24.4 | 30.3 | 38.4 |
| Percent male | 31% | 38% | 84% |
| Percent who have taught | 62% | 91% | 69% |
| Years R&D work experience | 0.2 | 1.4 | 10.1 |
| Nr. articles in professional journals | 0.1 | 0.7 | 2.6 |
| Nr. technical R&D reports | 0.0 | 0.2 | 8.8 |
| Nr. evaluation reports | 0.1 | 0.6 | 3.6 |
| Nr. chapters in professional books | 0.0 | 0.0 | 0.7 |
| Nr. other publications | 0.1 | 0.7 | 1.9 |
| Nr. tests/assessment instruments | 0.0 | 1.1 | 8.5 |
| Nr. programmed lessons | 0.2 | 0.9 | 8.1 |
| Nr. slide tapes/audio tapes | 0.0 | 0.5 | 2.1 |
| Nr. instructional films/T.V. | 0.0 | 0.6 | 0.3 |
| Nr. other R&D products | 0.0 | 2.3 | 2.3 |
| Percent with PhD's or EdD's | 0% | 14% | 37% |

## 2. RDD&E Activities Instrument

In the field test of this version of the self-rating instrument, we employed a simplified six level performance scale which omits all references to knowledge and focuses solely on performance.* The scale is as follows:

### Level of Performance

1. I cannot perform this activity even with supervision or guidance.
2. I can perform this activity if I have either detailed instruction or close supervision.
3. I can perform this activity if I am given enough general supervision or general instructions.
4. I can perform this activity quite satisfactorily without supervision or job aids.
5. I can perform this activity quickly and efficiently and can do a top quality job.
6. I consider myself an expert in this activity and can accomplish unusually difficult or completely novel work.

The RDD&E Activities Instrument consists of six scales, each containing ten items. The Activities Instrument, although shorter in length than the DD&E Diagnostic Instrument, covers a broader range of competencies including: (1) Research and Statistics, (2) Evaluation, (3) Instructional Product Development, (4) Planning and Technical Writing, (5) Dissemination, and (6) R&D Management.

---

*Please compare this performance scale with the proficiency scale used with DD&E Diagnostic Instrument (p.19 ). The first three levels of the proficiency scale (1, no knowledge; 2, don't really understand; 3, good understanding but haven't done it) have been combined into the rough equivalent of performance level 1 (can not perform this activity even with supervision or guidance). Proficiency scale level 4 corresponds approximately to performance scale level 2; 5 corresponds to 3; 6 corresponds to 4; and 7 corresponds to 5. Finally a sixth performance level ("expert") was added. Although the resulting 6 point performance scale avoids confounding knowledge with performance and may be more appropriate for use with advanced professionals, we consider the 7 point proficiency scale to be more useful for training program assessment.

Table B2 reports the item and scale means for the three groups;, and the P-levels based on one way analysis of variance tests of the difference among the means for the three R&D experience level groups. Table B3 reports the scale intercorrelations, means, standard deviations, and alpha coefficients for the six scales.

The data in Table B2 indicate that all six scales discriminate among the three groups with P-levels (substantially) less than .01 on all scales except Research and Statistics, where the P-level is less than .05. The student group with no R&D work experience (but with some training display average scale ratings close to or slightly above 3.0 on all scales ("can perform if given enough _general_ supervision or _general_ instructions").* By contrast, the high R&D experience group has average scale ratings above 4.0 ("I can perform this activity quite satisfactorily _without_ supervision or job aids") on four scales and above 3.7 on the other two (Research & Statistics, Dissemination). The group with modest R&D experience has scale average ratings that are intermediate on all six scales; however, the modest R&D experience group is only slightly higher than the no experience group on the Dissemination and the R&D Management scales.

It should be noted that the items appearing on the six self-rating scales have _not_ been selected on the basis of group discrimination; but rather for their coverage of RDD&E content that seemed appropriate for entry level professionals. Given very large differences in work experience and productivity indicated in Table 1, the general character of the self-rating results reported in Table 2 should be expected. For instance, note that the high R&D experience group has average ratings that are substantially higher than the other two groups on the R&D Management scale. In this case there are significant differences among the groups on all ten items on this scale. By contrast, significant

---

*Typically, the students who comprised this no R&D work experience group had completed one course in research design, one course in tests and measures, one course in instructional technology and one or two courses in statistics. About half had also completed a course in evaluation.

## TABLE B2. RDD&E ACTIVITIES ITEM MEANS

### 1. RESEARCH AND STATISTICS

| | R&D Experience | | | |
|---|---|---|---|---|
| | NONE | MODEST | HIGH | P-Level |
| 3. Choose (or design) appropriate statistical techniques for data analysis. | 3.08 | 2.81 | 3.53 | NS |
| 11. Use and interpret statistical regression techniques | 2.46 | 2.76 | 3.38 | NS |
| 12. Identify the types of questions which can and cannot be answered by different research designs. | 2.69 | 2.95 | 3.81 | .01 |
| 13. Design data collection procedures to maintain privacy or confidentiality in collecting, processing and storing information. | 3.08 | 3.86 | 3.97 | .03 |
| 19. Identify and articulate the problem in a research study. | 3.46 | 4.00 | 4.66 | .01 |
| 23. Formulate testable hypotheses in a research study. | 3.31 | 3.95 | 4.44 | .02 |
| 24. Prepare a coding scheme which will be used by a group of coders in preparing field test data for computer analysis. | 2.69 | 3.38 | 3.65 | NS |
| 27. Identify factors which can jeopardize internal and external validity. | 3.38 | 3.71 | 4.00 | NS |
| 33. Employ sampling theory and techniques to develop a stratified or clustered sampling plan. | 2.92 | 3.05 | 3.31 | NS |
| 42. Determine an experimental design and statistical techniques whose underlying theoretical assumptions are consistent. | 3.15 | 2.86 | 3.56 | NS |
| TOTALS | 30.2 | 33.3 | 38.3 | .05 |

TABLE B2 (Continued)

## 2. EVALUATION

| | R&D Experience | | | P-LEVEL |
|---|---|---|---|---|
| | NONE | MODEST | HIGH | |
| 1. Revise a test instrument on the basis of data collected during a pilot test of the instrument. | 3.00 | 3.71 | 4.41 | .01 |
| 5. Design a monitoring system that will provide data on the status of the operating system (such as actual vs. intended outcomes, unmet needs, problems, etc.). | 2.85 | 2.71 | 4.25 | .01 |
| 16. Analyze discrepancies between expected and actual test outcomes. | 3.31 | 3.33 | 3.91 | .05 |
| 18. Determine if new test instruments need to be developed by reviewing how well available test fit the evaluation objectives. | 3.08 | 3.57 | 4.06 | .01 |
| 21. Given the summary report of an initial tryout, suggest new methods to use in the second tryout to improve tryout validity. | 3.08 | 3.19 | 4.41 | .01 |
| 30. In planning the evaluation of a new training program, determine the types of subjects which should be used. | 3.38 | 3.71 | 4.28 | .02 |
| 44. Provided with product specifications, the test product, field test reports, and expert reviews make recommendations regarding the nature and extent of revisions required. | 3.15 | 3.81 | 4.28 | .01 |
| 49. Design data collection forms for a marketing study of an educational product. | 2.92 | 2.95 | 3.50 | NS |
| 50. Prepare a summary report of the test of your instructional product. | 3.31 | 3.67 | 4.34 | .01 |
| 57. Specify a set of procedures to be followed by evaluation personnel in administering test instruments. | 3.00 | 3.62 | 4.19 | .01 |
| TOTALS | 31.3 | 34.3 | 41.6 | .01 |

## TABLE B2 (continued)

### 3. INSTRUCTIONAL PRODUCT DEVELOPMENT

| | R&D Experience | | | |
|---|---|---|---|---|
| | NONE | MODEST | HIGH | P-LEVEL |
| 7. From a general description of a product's purposes, specify appropriate terminal objectives. | 3.69 | 4.14 | 4.34 | NS |
| 8. Prepare a sample of an educational product which can be presented to a specified target group in order to determine if the proposed product meets the group's needs. | 2.92 | 3.90 | 4.41 | .01 |
| 26. Given an educational product which you are developing, prepare tests and measures to assess student performance. | 2.84 | 3.29 | 3.38 | NS |
| 28. Convert a chapter in an instructional text into a branching programmed instruction sequence. | 3.38 | 3.71 | 4.28 | .02 |
| 34. With production personnel, determine the sequence of activities for production of instructional materials. | 3.08 | 3.76 | 4.53 | .01 |
| 40. Categorize instructional objectives in terms of a taxonomy (e.g., Bloom's Taxonomy of the cognitive domain or Gagne's conditions of learning). | 3.69 | 4.24 | 3.84 | NS |
| 46. Write exercises which the learner should do in order to master concepts or principles in an instructional unit. | 3.46 | 4.33 | 4.25 | .05 |
| 47. Evaluate alternative media and instructional methods to match them with instructional objectives and learner characteristics. | 3.23 | 3.90 | 4.12 | .05 |
| 58. Analyze job or task requirements to determine objectives for an instructional program or product. | 3.53 | 3.81 | 4.38 | .02 |
| 59. Determine questions you would like to have answered by the data collected in a tryout of your product. | 3.69 | 4.14 | 4.53 | .02 |
| TOTALS | 33.2 | 38.8 | 41.5 | .01 |

TABLE 32 (Continued)

## 4. PLANNING & TECHNICAL WRITING

| | R&D Experience | | | |
|---|---|---|---|---|
| | NONE | MODEST | HIGH | P-LEVEL |
| 4. Use bibliographic resources such as Research in Education and Current Index to Journals in Education to locate information supporting the need to develop an educational program or product. | 4.00 | 4.24 | 4.56 | NS |
| 9. Review first draft materials to determine if they comply with specifications regarding objectives, sequencing, and content. | 3.69 | 4.33 | 4.75 | .01 |
| 14. Conduct a search to determine if products exist which could meet your need(s) for in-materials. | 3.31 | 3.57 | 4.25 | .01 |
| 31. Write a detailed product description including all the necessary elements. | 3.15 | 3.48 | 4.28 | .01 |
| 32. Prepare a product design which is consistent with the stated overall purpose and use of the product. | 2.92 | 3.10 | 4.00 | .01 |
| 38. Locate existing methods or strategies which can be used in potential product development. | 3.46 | 3.57 | 4.25 | .01 |
| 39. Write scripts for instructional films, videotapes, or sound filmstrips. | 2.77 | 3.52 | 3.62 | NS |
| 51. In preparing a proposal for a development project, search for and select references which support the rational and technical approach of the project. | 3.46 | 3.81 | 4.66 | .01 |
| 52. Determine revisions that can be made within specified time constraints. | 3.15 | 3.38 | 4.19 | .01 |
| 53. Review the relevant research literature for evidence to support the selection of an instructional strategy. | 3.77 | 4.33 | 4.59 | .04 |
| TOTALS | 33.7 | 37.3 | 43.2 | .01 |

TABLE B2 (Continued)

## 5. DISSEMINATION

| | R&D Experience | | | P-LEVEL |
|---|---|---|---|---|
| | NONE | MODEST | HIGH | |
| 15. Confer with different types of users about their problems in the use of an educational product. | 3.62 | 4.09 | 4.50 | .01 |
| 25. Make arrangements to secure copyrights and copyright clearances where needed. | 3.07 | 2.62 | 2.94 | NS |
| 29. Determine the thoroughness of distribution which occurred in dissemination of an educational product. | 2.77 | 2.71 | 3.38 | .02 |
| 36. Deliver an oral presentation to a group of more than 20 professional educators to disseminate information about a project or product. | 3.08 | 3.76 | 4.53 | .01 |
| 37. Plan interviews with potential users for the purpose of determining a market for your product. | 3.46 | 3.57 | 4.16 | .02 |
| 41. Collect information on installation costs for your instructional product. | 3.08 | 2.81 | 3.56 | .02 |
| 43. Identify the crucial characteristics of a target group which may influence the dissemination effort. | 2.92 | 3.29 | 3.62 | NS |
| 54. Determine the effectiveness of using a demonstration to disseminate information to your target audience. | 3.15 | 3.10 | 3.97 | .01 |
| 56. Evaluate at least three different market research techniques applicable to a specific development. | 2.46 | 2.19 | 2.81 | .02 |
| 60. Obtain information about strategies used to market educational products similar to yours. | 3.31 | 3.43 | 3.69 | NS |
| TOTALS | 31.2 | 31.6 | 37.2 | .01 |

TABLE B2 (Continued)

### 6. R&D MANAGEMENT

| | R&D Experience | | | |
|---|---|---|---|---|
| | NONE | MODEST | HIGH | P-LEVEL |
| 2. After product installation, interview users to determine if adequate information was provided about how to install and use the product. | 3.62 | 3.86 | 4.44 | .02 |
| 6. Supervise professional R&D personnel. | 2.38 | 2.95 | 4.47 | .01 |
| 10. Derive a set of questions which could be sent with your product to reviewers in order to focus their review on critical aspects of your product. | 3.62 | 3.76 | 4.34 | .02 |
| 17. Write and submit proposals to obtain funding and to negotiate with funding agencies. | 2.92 | 2.48 | 4.38 | .01 |
| 20. Determine financial resources necessary to conduct a program or project and use accounting procedures to operate within a program or project budget. | 2.53 | 2.38 | 3.81 | .01 |
| 22. Given a problem statement, information on the history of the problem, objectives and possible solutions, write a 20-page proposal for solving the problem, including a rationale for the approach and a development schedule. | 2.92 | 3.00 | 4.38 | .01 |
| 35. Review a draft field test report to determine if it is correct, complete and ready for release. | 3.00 | 3.29 | 4.34 | .01 |
| 45. Prepare a memorandum which summarizes the actions taken and problems encountered in a meeting with your development team. | 3.84 | 4.38 | 4.81 | .01 |
| 48. Evaluate the feasibility and risks associated with alternative approaches for redesigning a product that has failed to meet specifications. | 2.92 | 2.90 | 3.81 | .01 |
| 55. Use management and planning systems such as PERT (Program Evaluation and Review Technique), PPBS (Program Planning Budgeting System), or Critical Path Analysis. | 2.76 | 2.38 | 3.16 | .02 |
| TOTALS | 30.5 | 31.4 | 41.9 | .01 |

group differences are least evident for the Research and Statistics scale, where six of the ten items display no significant differences.*

Turning now to Table 3, we note that there are moderately high correlations among the six scales with correlation coefficients ranging from .40 to .82. The Research and Statistics scale is (not surprisingly) strongly correlated (.79) with the Evaluation scale; but otherwise displays the lowest correlations with the remaining four scales (.40 to .51). Instructional Product Development is the next most independent scale with correlations ranging from .51 (with Research and Statistics) to .74 (with Planning and Technical Writing). By contrast, the Evaluation scale shows moderately high correlations with all five other scales. It's relatively high (.77) correlation with the Management scale can be easily understood when one examines the R&D Mangement scale item content (Table 32) and realizes that many of these items have evaluative elements. The Management scale, like the Evaluation scale, also displays moderately strong correlations with all the scales; the highest being .82 with Dissemination ** and the lowest being .51 with Research and Statistics.

### 3. Job Knowledge Questions

This instrument contains ten short answer essay questions. It has a one hour time limit. The ten questions involved the following. (The content of the following is sometimes abbreviated; the complete questions are listed in Appendix C.)

1. Identify important information missing in a product description.
2. Specify appropriate terminal objectives for a briefly described product.

---

*This result is possibly attributable to two factors. First, as is indicated by the standard deviations in Table 3, there is substantially greater variability for this scale. Second, our experienced R&D test subjects were recruited entirely from educational development agencies. The majority of these subjects were developers rather than researchers. Thus, when we checked the biographic data we found that, although the more experienced subjects tended to have taken more courses in research design and in statistics, there were no significant differences among the three groups in terms of the number of these courses taken.

**The unusually high correlation between Dissemination and Management is partly seen in the group means for these two scales. Generally neither the "No R&D Experience" nor the "Modest R&D Experience" groups rated themselves very high on either the Management or the Dissemination scales, but the "High R&D Experience" group generally rated themselves (relatively) much higher on both scales.

## TABLE B3

## RDD&E ACTIVITIES SCALE INTERCORRELATIONS, MEANS (M) STANDARD DEVIATIONS (SD) AND ALPHA INTERNAL CONSISTENCY COEFFICIENTS (A)

| Scale | M | SD | A | R&D 1 | EVAL 2 | DEV 3 | P&TW 4 | DIS 5 | MGT 6 |
|---|---|---|---|---|---|---|---|---|---|
| 1. Research and Statistics | 35.14 | 10.62 | .95 | - | .79 | .51 | .41 | .40 | .51 |
| 2. Evaluation | 37.26 | 8.31 | .93 | .79 | - | .71 | .66 | .68 | .77 |
| 3. Instructional Product Development | 39.00 | 8.36 | .91 | .51 | .71 | - | .74 | .64 | .58 |
| 4. Planning and Technical Writing | 39.44 | 8.17 | .93 | .41 | .66 | .74 | - | .78 | .75 |
| 5. Dissemination | 34.20 | 7.08 | .90 | .40 | .68 | .64 | .78 | - | .82 |
| 6. R&D Management | 36.33 | 8.48 | .91 | .51 | .77 | .58 | .75 | .82 | - |

3. Analyze and discuss a proposed product tryout in terms of the relevance and importance of the information it would yield.
4. List three of the basic criteria for judging the adequacy of revisions which have been made in [any] product.
5. Compare the advantages and disadvantages of sound movies and sound filmstrips as instructional media considered from the viewpoint of the developer, the distributor, and the user.
6. Indicate three questions a developer might want to answer in a tryout about student satisfaction with a programmed text.
7. Identify the information that could be obtained from a consultant review of a product that might not be obtained in a field test of the product.
8. List three questions (for each of three specified types of reviewers) which should be asked about a [specified] product.
9. Identify factors other than the product itself, which might account for difficulty subjects may encounter in using a [any] product effectively.
10. Outline the sequence of major steps that are most important in the production of a sound filmstrip between the two stages (a) definition of the instructional objectives and (b) collections of the first pilot test data.

Each item was scored three ways: (1) pass-fail (P-F), scored 1 or 2; (2) fraction (F) of essential elements covered in the response, scored 0-9; and (3) scope (SC) of application of knowledge going beyond essential requirements, scored 1-5*. The pass-fail (P-F) score was scored 2 (pass) if, and only if, all essential elements of the answer specified in the scoring manual were present (equal to a score of 9 on the 0-9 fraction score). It is not surprising then that the average intercorrelation of these two scores, averaged over all ten items (with averages based on Z-transformations) is .87. The scope score, which may be considered a "bonus" score for responses which reveal a "scope" of application of knowledge going beyond the specific essential requirements set for the question, was scored 1-5. Conceptually, the scope score is considered to be independent of the P-F and fraction score. A student could fail the item and still receive a 2, 3, 4, or 5 on scope, although one might expect that more knowledgeable subjects would receive higher marks on all three scores. The average intercorrelation of the scope score with the fraction score was .48; with the P-F score it was .44. These correlations are positive and significantly different from zero, but they do indicate that the scope score does contain information not present in the other two scores. Although there are significant intercorrelations among the three scores based on the same questions, the intercorrelations among the ten questions were surprisingly small, with only 16% showing correlation values above .25 ($P < .05$). A factor analysis of the 30 scores revealed that there were perhaps as many factors as there were questions. Apparently the knowledge required to deal with these ten Job Knowledge Questions is moderately specific to each item. There are modest relations, for instance, among the several questions dealing with review or tryout, and there is a tendency for fraction or scope scores of one question to be related to scope scores of other questions, but none of these relationships are particularly strong.

---

*See Appendix D for complete scoring instructions.

We now turn to the question of whether the Job Knowledge Questions scores are capable of discriminating among the three R&D work experience groups. On examining the item means for the three groups we noted that both the fraction and the scope score means were higher for the two groups with R&D work experience than for the group with no R&D experience on all ten questions. But analysis of variance indicated that only four of the ten questions had score mean differences that were significant at the .05 level. However the total scores (summed over the 10 items) are all significant. See Table B4.

Note that the modest R&D experience group has slightly higher total scores than the high experience group (all non-significant). Both of these experienced groups have higher scores than the no experience group.

The pass-fail scores are troubling, primarily because they indicate that even the more experienced groups were averaging only 6.9 and 6.7 questions "passed" out of the ten. We have two conjectures; first, the one hour time limit is probably too short to permit a knowledgeable subject to answer all ten questions; second, the scoring standards are perhaps unrealistically high. We would recommend that the user adapt this test possibly by both reducing the number of questions and by relaxing the pass-fail criteria.

The fraction scores are based on a 0-9 scale for each item which translates to a 0-90 possible range for the totals. Table B4 indicates (in parentheses) what the F scores mean in terms of percentage of the possible score of 90,

## TABLE B4

## MEANS, F-RATIOS AND P-LEVELS FOR THE F, P-F, AND SC TOTAL SCORES ON JOB KNOWLEDGE QUESTIONS

| | R&D EXPERIENCE GROUP | | | F-Ratio | P-Level |
|---|---|---|---|---|---|
| | NO | MODEST | HIGH | | |
| Pass-Fail (P-F)<br>(P-F as % possible) | 15.3<br>(53%) | 16.9<br>(69%) | 16.7<br>(67%) | 3.81 | 0.027 |
| Fraction (F)<br>(F as % possible) | 70.6<br>(78%) | 79.5<br>(88%) | 77.9<br>(87%) | 4.89 | 0.011 |
| Scope (SC)<br>(SC as % possible | 25.2<br>(50%) | 31.7<br>(63%) | 29.8<br>(60%) | 7.54 | 0.002 |

namely that students with no R&D work experience averaged 78% while the more experienced subjects averaged approximately ten percentage points higher. The scope scores also indicate marked differences between the experienced and inexperienced groups in terms of their ability to deal with the questions in ways that apply knowledge going beyond the pre-established minimum requirements.

Summary.

The Job Knowledge Questions test does discriminate between those with no R&D work experience and those with some experience. The differences are not large, but they are statistically significant. However, there is no evidence that those with substantially more experience perform better on this test, in fact the scores of the high experience group are slightly lower than those of the modest experience group. The three total scores have substantial intercorrelation (.88 between F and P-F; .73 between F and SC; and .71 between P-F and SC). Intercorrelation of scores based on each item indicate that the ten items are not highly intercorrelated, indeed only 16% of the between questions score intercorrelations were significant and none exceeded .50.

### 4. An Exercise in Educational Development

This simulation test consists of four related tasks. Task 1 is concerned with preparing guidelines for the development of a simple product (a section of a toy librarian's manual for parent training). Task 2 deals with preparing for the tryout of the parent training resources kit. Task 3 requires responses to the tryout evaluation report by critiquing the tryout and recommending revisions in the product and in the tryout procedure. Task 4 requires a rough outline for a sound filmstrip script. Following the same rationale used in scoring the Job Knowledge Questions, each task was scored for pass-fail (P-F), fraction (F) and scope (SC) according to instructions contained in the scoring manual. Each task was also scored in terms of the extent to which the simulation directions (DIR) were followed. Hence there are four scores that are of the same type for each task (P-F, F, SC and DIR). In addition to these 16 scores, tasks are also scored on a total of 11 specific elements (e.g., quality of writing of the guideline produced in Task 1, the quality of the tryout questions listed in Task 2, the suggested revisions of Task 3, the handling of visual elements in the script outline of Task 4).

Simulation score relationships.

The 27 scores of the simulation exercise were intercorrelated and then factor analyzed. As had been the case in the knowledge test, there were moderate to substantial intercorrelations among the several scores based on the same task, but smaller correlations among scores for different tasks. However, in contrast to the small proportion of significant intercorrelations found in the knowledge test, fully one half of the between tasks correlations were significant in the simulation test. A factor analysis indicated that the first principal axis factor accounted for 34% of the trace (total covariance) and that six factors, accounting for a total of 74% of the trace, had eigenroots above 1.0. These results indicate that the simulation test is factorially complex, but that there is a modest "general competence" factor that runs through it.*

Every one of the 27 scores displays a significant loading on (i.e., correlation with) the first principal axis; however these loadings range from .33 to .81. Thus, the scores exhibit varying degrees of association with this general performance factor. When the six factors are rotated to a varimax solution, (see Table B5), we discover that every score in Task 1 displays a significant loading on Factor 1. (However, the Task 1 scope score and the guideline instructions score also show substantial loadings on Factor 6, and Task 1 P-F score has a modest loading on Factor 2.) The only score loading significantly on Factor 1 and not in Task 1 is the scope score on Task 2. Hence, Factor 1 is almost exclusively identified with Task 1 (Preparing Guidelines).

The scores on Task 2 (Preparations for Tryout) are split primarily between Factors 2 and 3; but the scope score on Task 2 has modest loadings on four of the six factors. Factor 2 is best identified by the "specific" score for specification of test sites and subjects. Factor 3 is identified by the "specific" score

---

*This factor analysis contrasts with the knowledge test where only 18% of the trace was extracted by the first principal axis, and where ten factors, accounting for 81% of the trace, had eigenroots above 1.0. Eight factors extracted 73% of the trace. Hence the knowledge test required two more factors than the simulation test to extract the same percentage of the trace, and its first principal axis extracted a much smaller percentage of the trace. In other words, the knowledge test has even greater factorial complexity and has a much smaller "general factor" (18% as opposed to 34% extracted on the first principal axis).

## Table B5

## Factor Loadings, Score Means and Standard Deviation (SD) for 27 Simulation Scores

| Task | Score | FACTOR | | | | | | SCORE | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | Mean | S.D. |
| 1 | Pass Fail | .69 | .28 | -.10 | -.18 | .06 | .16 | 1.21 | 0.41 |
| 1 | Fraction | .93 | .09 | .16 | .08 | .07 | .18 | 6.53 | 1.79 |
| 1 | Scope | .41 | .20 | .10 | .02 | .13 | .35 | 3.35 | 1.17 |
| 1 | Following Task 1 Directions | .79 | -.03 | .10 | .21 | -.09 | -.12 | 7.89 | 1.72 |
| 1 | Consistency with Information Provided | .83 | .12 | .11 | -.01 | .15 | -.09 | 5.26 | 2.05 |
| 1 | Specification of Instructional Objective | .73 | .24 | .05 | .12 | .18 | .05 | 5.77 | 3.20 |
| 1 | Guideline Instructions | .38 | -.05 | .24 | -.02 | .10 | .58 | 8.08 | 1.73 |
| 1 | Guideline Writing Quality | .66 | -.16 | .16 | .13 | .13 | .19 | 7.11 | 0.86 |
| 2 | Pass Fail | .16 | .87 | .02 | .20 | .20 | .03 | 1.50 | 0.50 |
| 2 | Fraction | .23 | .67 | .61 | .21 | .20 | .03 | 8.29 | 0.98 |
| 2 | Scope | .34 | .45 | .26 | .08 | .40 | -.02 | 3.38 | 0.90 |
| 2 | Specification of Test Sites and Subjects | .11 | .86 | .19 | .24 | .07 | .02 | 7.00 | 1.95 |
| 2 | Following Task 2 Directions | .12 | .09 | .91 | .16 | .06 | .19 | 8.93 | 0.49 |
| 2 | List of Tryout Questions | .23 | .15 | .85 | -.03 | .22 | .04 | 8.77 | 0.75 |
| 3 | Following Task 3 Directions | .12 | .09 | .91 | .15 | .06 | .19 | 8.88 | 0.98 |
| 3 | Revisions Recommended | -.01 | .01 | .44 | .36 | .09 | .59 | 8.35 | 1.73 |
| 3 | Changes Recommended in Tryout | -.06 | -.16 | .47 | .64 | .04 | -.17 | 8.17 | 2.03 |
| 3 | Critique of Tryout Method | .17 | .20 | .09 | .86 | .26 | .02 | 6.18 | 2.20 |
| 3 | Pass Fail | .06 | .41 | -.17 | .72 | .05 | .18 | 1.24 | 0.43 |
| 3 | Fraction | .08 | .09 | .33 | .86 | .22 | .17 | 7.29 | 1.66 |
| 3 | Scope | .00 | .24 | .15 | .73 | .31 | .01 | 3.42 | 1.00 |
| 4 | Pass Fail | .07 | .23 | -.12 | .35 | .60 | .13 | 1.30 | 0.46 |
| 4 | Fraction | .15 | .16 | .24 | .19 | .89 | .14 | 7.36 | 1.81 |
| 4 | Scope | .16 | .12 | .04 | .25 | .71 | .25 | 3.12 | 1.19 |
| 4 | Following Task 4 Directions | .02 | -.01 | .00 | -.02 | .75 | .15 | 8.47 | 1.61 |
| 4 | Visual Element of AV Scropt | .08 | .21 | .13 | .20 | .79 | .39 | 6.68 | 2.15 |
| 4 | Quality of Script Outline | .19 | .00 | .38 | .19 | .65 | -.32 | 7.39 | 2.15 |
| | PERCENT OF TRACE | .16 | .11 | .14 | .14 | .14 | .05 | — | — |

score for the list of tryout questions in Task 2, by the following directions scores of Tasks 2 and 3*, and by the revisions and changes recommended in Task 3.

The scores on Task 3 are even more factorially complex since they display substantial loadings on Factors 2, 3, 4, 5 and 6; however, Factor 4 is most strongly associated with Task 3, since the P-F, F, SC, critique of tryout method, and changes recommended in tryout all show strong loadings on Factor 4. Similarly, Task 4 has loadings on several factors, but it is primarily associated with Factor 5. Factor 6, a much weaker factor, accounting for only five percent of the trace, is associated with several specific scores on different tasks, most notably: the quality of instructions provided in the guidelines of Task 1, the handling of suggested revisions based on field test results of Task 3, and the handling of visual elements and the quality of script (negative loading!) in Task 4. The scope scores of Tasks 1 and 4 also had smaller but significant loadings on the sixth factor. Note also that all four of the scope scores display loadings of .25 or more on two , and sometimes three or four, factors; which tends to confirm the character of this type of score as the ability to apply a broad range or "scope" of knowledge or skill.

To summarize, the factor analysis indicates that there are six factors associated with the simulation test. In terms of size of loadings, Task 1 is associated primarily with one factor (Nr 1) and to a lesser extent Task 4 is primarily associated with one factor (Nr 5). However the scores in Tasks 2 and 3 load on several different factors. We conclude that the simulation represents a complex of several different performance dimensions.

Turning to the means in Table B4 note that the decimal part of the pass-fail score means is equal to the percent passing, thus only 21, 50, 24, and 30 percent of the subjects "passed" Tasks 1, 2, 3, and 4 respectively. In retrospect, the pass-fail standards appear to be too strict. The fraction means are 6.53, 8.29, 7.29 and 7.36 on the 0-9 point scale. And all scope means are above 3 on the 1-5 point scale. All eleven "specific" items and the following directions and fractions items were all scored 0 to 9. Subjects did quite well on

---

*The very high loadings for following directions in Tasks 2 and 3 are probably spurious since they can be attributed to extremely low scores made on both tasks by one or two persons.

Table B6

Means for Three R&D Experience Groups, F-Ratios and P-Levels For Tests of Differences on Simulation Test Scores

| Task | Score | R&D Experience | | | Group Means | |
|---|---|---|---|---|---|---|
| | | No Exp | Modest Exp | High Exp | F-Ratio | P-Level |
| 1 | Pass Fail (Guidelines) | 1.15 | 1.29 | 1.19 | 0.51 | NS |
| 1 | Fraction | 5.38 | 6.76 | 6.84 | 3.55 | .03 |
| 1 | Scope | 2.69 | 3.43 | 3.56 | 2.71 | .07 |
| 1 | Follow Directions | 7.00 | 7.86 | 8.22 | 2.37 | .10 |
| 1 | Consistency | 4.31 | 5.67 | 5.38 | 1.89 | NS |
| 1 | Specification of Objective | 4.23 | 6.90 | 5.66 | 2.96 | .06 |
| 1 | Guideline Instructions | 7.54 | 8.24 | 8.19 | 0.77 | NS |
| 1 | Guideline Writing Quality | 6.69 | 7.00 | 7.34 | 3.06 | .05 |
| 2 | Pass Fail (Tryout) | 1.08 | 1.76 | 1.50 | 9.32 | .001 |
| 2 | Fraction | 7.54 | 8.57 | 8.41 | 5.50 | .01 |
| 2 | Scope | 2.54 | 3.86 | 3.41 | 11.12 | .001 |
| 2 | Specification of Test Sites & Subjects | 5.38 | 7.81 | 7.12 | 7.45 | .002 |
| 2 | Following Task 2 Directions | 8.69 | 9.00 | 9.00 | - | NS |
| 2 | List of Tryout Questions | 8.46 | 8.56 | 8.84 | 1.37 | NS |
| 3 | Pass Fail (Revisions) | 1.08 | 1.29 | 1.28 | 1.20 | NS |
| 3 | Fraction | 6.31 | 7.38 | 7.62 | 3.11 | .05 |
| 3 | Scope | 2.62 | 3.52 | 3.69 | 6.24 | .004 |
| 3 | Following Task 3 Directions | 8.38 | 9.00 | 9.00 | - | NS |
| 3 | Revisions Recommended | 7.69 | 8.62 | 8.44 | 1.23 | NS |
| 3 | Critique of Tryout Method | 5.00 | 6.43 | 6.50 | 2.40 | .10 |
| 3 | Changes Recommended | 7.46 | 7.90 | 8.62 | 1.80 | NS |
| 4 | Pass Fail (Script) | 1.08 | 1.38 | 1.34 | 2.03 | NS |
| 4 | Fraction | 6.54 | 7.38 | 7.69 | 1.88 | NS |
| 4 | Scope | 2.54 | 3.14 | 3.34 | 2.17 | NS |
| 4 | Following Task 4 Directions | 8.38 | 8.48 | 8.50 | 0.02 | NS |
| 4 | Visual Elements,AV Script | 5.92 | 6.57 | 7.06 | 1.34 | NS |
| 4 | Quality of Script Outline | 6.54 | 7.57 | 7.62 | 1.27 | NS |
| Tot | Pass Fail | 4.38 | 5.71 | 5.31 | 5.38 | .007 |
| Tot | Fraction | 25.77 | 30.10 | 30.56 | 5.90 | .005 |
| Tot | Scope | 10.38 | 13.95 | 14.00 | 9.17 | .001 |
| Tot | Following Directions | 32.46 | 34.33 | 34.72 | 3.17 | .05 |
| Tot | Consistency with Simulation Information* | 23.92 | 29.67 | 28.56 | 7.03 | .002 |

---

*The Consistency Total Score is the sum of the following four task scores: Consistency in Task 1, Specification of Test Sites and Subjects in Task 2, Revisions Recommended in Task 3, and Quality of Script Outline in Task 4. These four task scores best reflect the test subject's ability to cope with and respond consistently with the information (not directions) provided.

following directions with means of 7.89 to 8.93. "Specific" scores with relatively high means(8.0 or higher) are: guideline instructions of Task 1, list of tryout questions of Task 2, and revisions recommended and changes recommended both in Task 3. Relatively poorer performance is noted in the case of consistency (of product guideline) with information provided and specification of instructional objectives, both in Task 1, critique of tryout method in Task 3, and handling of visual elements in the script outline of Task 4.

What about differences among subjects in terms of their R&D experience? Turning to Table 6, we see that scores on the four tasks varied in their discrimination of the three R&D experience groups. We note first that the no R&D experience group has consistently lower means than either of the two experienced groups on all 27 scores and on all the total scores. The means for the modest experience and the high experience groups are fairly close to each other. The differences between groups on the fraction and scope scores in Tasks 1, 2, and 3 are all significant or near significant. In addition, the differences for guideline writing quality (Task 1) and specification of test sites and subjects (Task 2) are both significant. By contrast, none of the group differences for scores on Task 4 are significant. (Which may well suggest that Task 4 could be easily omitted from the battery, except possibly in cases where specific training in audio~~visual~~ production is part of the program of instruction.)

While not all of the individual item scores show significant differences, all of the total scores do. Perhaps most notable is the total for scope which has a P-level of less than .001. There are no appreciable differences between the modest and the high experience groups on the means for totals, but both of the experienced groups have means well above the no experience group on all five totals: pass-fail, fraction, scope, following directions and consistency with simulation information.

### 5. Relationships Between Self-Ratings, Knowledge, and Performance

Table 7 reports the intercorrelations among the total scales for the self-ratings, the knowledge test, and the simulation test. As we have noted in previous sections, all the total scores within the same instrument are significantly intercorrelated. However, our interest in Table 7 is in the correlation

## TABLE B7

### Intercorrelations of Self-Ratings, Knowledge and Simulation Total Scales

(N=66) (decimals omitted)

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Self Rating Research | - | 79 | 51 | 41 | 40 | 51 | 12 | 08 | 09 | 05 | 11 | 04 | -04 | 07 |
| 2 | Self Rating Evaluation | | - | 71 | 66 | 68 | 77 | 25 | 23 | 26 | 14 | 13 | 14 | 04 | 12 |
| 3 | Self Rating Product Development | | | - | 74 | 64 | 58 | 38 | 25 | 41 | 12 | 13 | 19 | 02 | 11 |
| 4 | Self Rating Technical Writing | | | | - | 78 | 75 | 22 | 18 | 31 | 15 | 01 | 14 | 15 | 06 |
| 5 | Self Rating Dissemination | | | | | - | 82 | 11 | 10 | 17 | 02 | -16 | 06 | 03 | -11 |
| 6 | Self Rating R&D Management | | | | | | - | 16 | 17 | 21 | 18 | -05 | 14 | 15 | 08 |
| 7 | Knowledge Fraction | | | | | | | - | 88 | 73 | 42 | 40 | 40 | 32 | 46 |
| 8 | Knowledge Pass-Fail | | | | | | | | - | 71 | 38 | 39 | 34 | 25 | 40 |
| 9 | Knowledge Scope | | | | | | | | | - | 38 | 39 | 45 | 31 | 43 |
| 10 | Simulation Fraction | | | | | | | | | | - | 68 | 76 | 79 | 88 |
| 11 | Simulation Pass-Fail | | | | | | | | | | | - | 64 | 31 | 64 |
| 12 | Simulation Scope | | | | | | | | | | | | - | 54 | 69 |
| 13 | Simulation Following Directions | | | | | | | | | | | | | - | 73 |
| 14 | Simulation Consistency | | | | | | | | | | | | | | - |

Correlations greater than .24 are significant at .05 level.
Correlations greater than .32 are significant at .01 level.

between scores on the three different instruments. The results can be summarized briefly as follows: (1) all of the simulation and the knowledge total scores are significantly correlated (ranging from .31 to .46), (2) none of the self-rating scales are significantly correlated with the simulation test, (3) some of the knowledge test and the self-rating scales are significantly correlated. Knowledge scope correlates .26 with the evaluation self-ratings, .41 with the product development self-ratings, and .31 with the technical writing self-ratings. The knowledge fraction score and the knowledge pass-fail score also display very modest correlations with these three self-ratings, with three of the six correlations reaching values significant at the .05 level. It is significant that the context of the knowledge test items deals most directly with product development, product evaluation and, to a lesser extent, technical writing, but does not deal with research, dissemination or R&D management. The pattern of significant correlations between the knowledge test totals and three of the self-rating scales (but not with the other three self-rating scales) is as would be anticipated.

Hence the data in Table 7 indicate that the knowledge test scores are significantly related to simulation scores and to self-ratings based on comparable content areas. The unanticipated result in Table 7 is the fact that none of the self-rating total scores is significantly related to any of the simulation scores. Given the clear ability of the self-ratings to discriminate among groups with known differences in R&D experience, and given the fact that (despite the relatively high intercorrelations among the six self-rating scales) the self-rating totals display a pattern of correlations whose magnitude and significance is consistent with the content of the knowledge test, we can only conclude that the self-rating scales are valid but that the three types of instruments represent increasing degrees of competence specificity with the self-ratings the most general, the knowledge test intermediate, and the simulation test the most specific. Some evidence supporting this explanation is found when we examine the relations between Biographic Information and the total scores for the three types of instruments. The self-rating scales are significantly related to membership in professional associations, authorship of journal articles, numbers of tests, R&D technical reports, and other publications. However, only the knowledge test scope score shows significant correlations with any of the ten R&D productivity measures; and in this case there were only two barely significant correlations, one with number of evaluation reports and the other with number of "other" products. None of the simulation totals display significant correlations with any of the R&D productivity measures.

Turning to the Biographic Information data on numbers of courses taken, we found a similar pattern. The largest number of and strongest correlations are found between the self-rating scales and courses taken, with substantially fewer significant correlations between number of courses taken and the knowledge test or the simulation test. As we might expect, there are modest to substantial correlations (.41 to .63) between both the Research & Statistics and the Evaluation rating totals and the numbers of courses taken in research design, in tests and measures, and in statistics, as well as "other" (than educational) psychology courses. The Instructional Product Development rating scale is only correlated with number of tests and measurement courses (.27) and number of statistics courses (.26); and the Dissemination Scale is only related to number of research design courses (.29); however, the Technical Writing Scale is significantly related to number of Communication Theory courses (.26), number of Test and Measurement courses (.33), number of other Psychology courses (.25) and number of other Social Science courses (.26). The Management Scale mirrors the significant patterns found for the Research and Statistics and for the Evaluation scales; correlating with Research Design courses (.38), Tests and Measurement courses (.42), Statistics courses (.31), and other Psychology courses (.41).

In the case of the Knowledge Test, all three totals (F, P-F, SC) correlate (.27 to .30) with number of courses in Instructional Technology. The only other significant correlations are: .243 for knowledge P-F and number of Creative Writing courses, and .26 for knowledge scope and number of courses in Philosophy and Humanities.

There are only four barely significant correlations between courses and the simulation totals, namely: Simulation fraction score with number of other Social Science courses (.25); simulation P-F with Journalism courses (-.27) and simulation directions with Philosophy and Humanities courses (.26).

There are an average of 5 significant correlations between number of courses and any one of the self-rating scales, less than two significant correlations between courses and any one of the knowledge test scores, and less than one significant correlation per simulation score. Among the ten measures of R&D productivity, we find an average of over four significant correlations per self-rating scale, less than one significant correlation per knowledge total, and no significant correlations with the simulation totals.

## 6. Summary

Given the above field test results, we conclude that the Educational Product Development Battery provides three different kinds of valid information on DD&E competencies. The self-rating instrument provides quite broad but relatively superficial information on perceived competence in six areas: Research and Statistics, Evaluation, Instructional Product Development, Technical Writing, Dissemination, and R&D Management. Although moderately to highly intercorrelated, the six scales provide significant and meaningful differentiations among subjects with no R&D work experience, subjects with modest experience, and subjects with much experience. Moreover the self-rating scales display significant and meaningful correlations with numbers and types of courses taken and with numbers and types of R&D products authored or produced. The three self-rating scales with content most similar to the content of the knowledge test (development, evaluation, and technical writings) are significantly correlated with the knowledge test (Research & Statistics, Dissemination, R&D Management). There are no significant correlations between the self-rating scales and the simulation scale totals. This absence of significant correlations was not anticipated, but may be attributable to the simulation test covering a far more specific area of competence content than the self-rating instrument.

The knowledge test covers a much narrower range of content than the self-rating scales, yet it is broader in scope than the simulation. While only a modest number of the scores based on an individual question significantly discriminate among the three groups with different levels of R&D experience,differences on all three of the total scores are highly significant. The knowledge test does not discriminate between those with high and those with modest R&D work experience but it does discriminate between those with experience and those who have none. Knowledge Test scores are significantly but very modestly related to the self-rating scales with corresponding general content (Instructional Product Development, Evaluation, Technical Writing), and to number of courses taken in instructional technology. The Knowledge Scope score is also significantly related to number of evaluation reports authored and number of "other" products produced. The Knowledge Test (which was designed to include content roughly paralleling that of the Simulation Test) is significantly but only modestly correlated (.25 to .46) with every one of the Simulation Test total scores.

All the Simulation Test total scores (in addition to their significant correlation with the knowledge test totals) discriminate significantly between those with no R&D work experience and those with R&D work experience. However, as in the case of the Knowledge Test, the Simulation Test was designed to appraise "entry level" competence and not higher levels. Neither instrument discriminates between subjects with relatively modest R&D experience (average of 1.4 years) and those with much greater R&D experience (average 10.1 years). Perhaps because the Simulation Test deals with a much narrower range of competencies, it displays no significant correlations with any of the six self-rating scales, or with any of the ten measures of R&D productivity.

Each of the three instruments tends to provide different kinds and amounts of information. The lack of strong correlations among the three types of instruments indicated that one instrument can not be substituted for another.* Moreover, aside from the moderate to strong intercorrelations among the six self-rating scales, the correlations between knowledge questions or simulation tasks within instruments are not strong; hence there is substantial specificity among the ten knowledge questions and the four simulation tasks.

---

*Using No R&D Work Experience (=0) vs Modest or High R&D Work Experience (=1) as a dummy criterion, and employing the total scores as predictors we obtained the following multiple correlations: .46 for the six Self-Rating scales; .42 for the three Knowledge Test Scales; .50 for the five Simulation Scales; .53 for Self-Ratings and Knowledge Test, .53 for Knowledge Test and Simulation Test, .61 for Self-Ratings and Simulation, and .62 for all three instruments. All of these multiple correlations are significantly different from zero correlation. Note that the Self-Ratings plus the Simulation produces almost as high a multiple correlation as all three instruments combined. From this point of view the Knowledge Test provides very little additional predictive information beyond that supplied in the Self-Ratings and the Simulation. But note carefully that these results are relevant to this one criterion.

From the standpoint of scores attained, all three types of instruments discriminate at statistically significant levels between those with no R&D work experience and those with some R&D work experience. However, highly experienced persons do not necessarily do better than less experienced persons on either the knowledge or the simulation tests. Only the self-rating scales, especially Dissemination and R&D Management, separate the highly experienced from those with more modest R&D work experience.

In terms of scale differences, those with no experience tend to rate themselves about one performance level lower than the highly experienced ("can perform if given enough general supervision or general instructions" vs "can perform quite satisfactorily without supervision or job aids") about ten percentage points lower on the knowledge test (78% vs 87% of possible score on fraction score and 63% vs 74% on scope score), and thirteen percentage points lower on the simulation test fraction score totals (72% vs 85% of possible score) and eighteen percentage points lower on the simulation test scope score totals (52% vs 70%).

In retrospect, we wish that we had included a group of students who had just entered the DD&E program and were without significant relevant training since we expect that the differences observed for such a group would be even larger. Please recall that even our No R&D Work Experience group had taken courses in Research Design, Tests and Measurement, Statistics, Instructional Technology and Evaluation. Hence their self-appraisal, knowledge, and performance skills would be expected to be substantially above those with neither relevant academic training nor relevant R&D work experience. However, our primary intent was to "calibrate" the battery against advanced students and experienced developers, since it is unlikely that the battery would ever be used in practice with entering students.

Given the results reported, we recommend that simulation Task 4 be deleted since it is the only task which showed no significant discrimination between the R&D experience groups. This should reduce the simulation to approximately three hours duration. Second, we recommend deleting at least two or three knowledge questions or alternatively extending the one hour time limit on the knowledge test. The one hour time limit for ten essay questions is unreasonable. It tends to place too much emphasis on "speed" and too little on "power." Finally, we urge the user to consider lowering the pass-fail criteria which are indicated for the knowledge questions and the simulation tasks. Although these criteria seemed to be reasonable at the time they are unrealistically high in terms of the performance of experienced subjects. *

---

*The pass-fail criteria which were applied in this field test required that the subject's response include all or virtually all of a number of elements. In most cases, the subject "failed" if even one element was missing.

## REFERENCES

AERA Task Force See Worthen, Anderson, and Byers, 1971 and specific papers. 1970-1971.

Anderson, R. D., Soptick, J. M., Rogers, W. T., and Worthen, B. R. *An Analysis and Interpretation of Tasks and Competences Required of Personnel Conducting Exemplary Research and Research-Related Activities in Education*. AERA Task Force Technical Paper No. 23. Washington, D. C.: American Educational Research Association, 1971.

Brown, Rex V. *Research and the Credibility of Estimates: An Appraisal Tool for Executives and Researchers*. Boston, Massachusetts, Harvard University, 1969.

Campbell, D. T., & Fiske, D. W. *Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix*. Psychological Bulletin, 1959, *56*, 81-105.

Clark, D. L. and Hopkins, J. E. *A Report on Educational Research, Development, and Diffusion Manpower, 1964-1974*. Bloomington, Indiana: Indiana University Research Foundation, 1969.

Ebel, R. *Some Limitations of Criterion-Referenced Measurement*. Paper prepared for AERA Symposium "Criterion-Referenced Measurement: Emerging Issues," Minneapolis, March 1970.

Elam, S. *Performance-Based Education: What is the State of the Art*? Washington, D. C.: American Association of Colleges for Teacher Education, 1971.

Ferguson, R. L., Novick, M. R. *Implementation of A Bayesian System For Decision Analysis in A Program of Individually Prescribed Instruction*. Iowa City, Iowa, Research and Development Division, 1973.

Hood, P. D., Banathy, B., Ward, J., and others. Design of a Functional Competence Training Program for Development, Dissemination, and Evaluation Personnel at Professional and Para-Professional Levels in Education. Far West Laboratory, Berkeley, California. December, 1970.

Hood, P. D., Banathy, B. Design of a Functional Competence Training Program for Development, Dissemination, and Evaluation Personnel at Professional and Para-Professional Levels in Education. Interim Report, Far West Laboratory, Berkeley, CA April, 1973.

Hood, P. D., Ha assy, B., Lash, A., and Ward, J. *Development of Assessment Instruments for Competence-Based Education: The Educational Development, Diffusion and Evaluation Training Program*. San Francisco, CA: Far West Laboratory for Educational Research and Development, November, 1973.

Hood, P. D. and Lash, A. *A Report on the Field Test and Revision of the DD&E Diagnostic Instrument*. San Francisco, DA: Far West Laboratory for Educational Research and Development, November, 1974.

Hood, P. D., Lash, A., and Blackwell, L. *Development of Assessment Instruments for Competence-Based Education: The Educational Development, Diffusion and Evaluation Training Program--Technical Plan for Pilot Validation of A Development Competence Assessment Battery*. San Francisco, CA: Far West Laboratory for Educational Research and Development, November 1974.

Houston, W. R. and Howsam, R. B. (Eds.) *Competency-based Teacher Education Progress, Problems and Prospects*. Palo Alto, CA: Science Research Associates, 1972.

Jackson, R. *Developing Criterion-Referenced Tests*. Educational Testing Service, ERIC Clearinghouse on Tests, Measurement, & Evaluation, Princeton, N. J., 1971.

Merwin, J. C. *Performance-Based Teacher Education: Some Measurement and Decision-Making Considerations*. Washington, D.C.: American Association of Colleges for Teacher Education, 1973.

Popham, W. J. (Ed.) *Criterion-Referenced Measurement: An Introduction*. Englewood Cliffs, N. J.: Educational Technology Press, 1971.

Popham, J. W., Cary, G. R., Chilstrom, B., Leps, A. L., Miller, T. M., and Saxe, R. "Performance Tests for Instructional Developers: An Exploratory Investigation." *Educational Technology*, March, 1974 pp. 57-61.

Sanders, J. R., & Worthen, B. R. An Analysis of Employers' Perceptions of the Relative Importance of Selected Research and Research-Related Competences and Shortages of Personnel with Such Competences. (AERA Task Force Technical Paper No. 3). Washington, D. C.: American Educational Research Association, 1970.

Schalock, H. D., Thomas, G. P., Morse, K. A., Smith, C. A., & Ammerman, H. L. *The Oregon Studies in Educational Research, Development, Diffusion, and Evaluation, Volume I, Summary Report*. Monmouth, Oregon: Teaching Research, 1972.

Schalock, H. D., and Thomas, G. P. *Some Observations on the Contribution of the Oregon Studies in Educational RDD&E to the Development of Competency Based Personnel Development Programs*. Monmouth, Oregon: Teaching Research, 1973.

Swaminathan, H., Hambleton, R. K., & Algina, J. *A Bayesian Decision-Theoretic Procedure for Use with Criterion-Referenced Tests*. *Journal of Educational Measurement*, 1975, *12*, 87-97.

Turner, R. L. *Levels of Criteria*. Appendix A, *The Power of Competency-Based Teacher Education*, edited by Benjamin Rosner and others. Princeton, N. J.: Educational Testing Service; Washington, D. C.: U.S. Department of Health, Education and Welfare, Office of Education, 1971.

Ward, J. *Development and Assessment of Caucational DD&E Competence Assessment Instruments*. San Francisco: The Far West Laboratory for Educational Research and Development, 1973.

Worthen B. R., & Byers, M. L. *An Exploratory Study of Selected Variables Related to the Training and Careers of Educational Research and Research-Related Personnel*. Washington, D. C.: American Educational Research Association, 1973.

Worthen, B. R. Anderson, R. D., and Byers M. L. *A Study of Selected Factors Related to the Training of Researchers, Developers, Diffusers and Evaluators in Education*. Final Report. Washington, D. C.: American Educational Research Association, 1971.

Worthen, B. R., Millman, J., Page, E. B., & Brezezinski, E. J. Development of a Pilot Test of Selected Competences in Educational Research, Evaluation, Development and Diffusion. (AERA Task Force Technical Paper No. 29). Washington, D. C.: American Educational Research Association, 1971.

# Appendix C

## Instructional Product Development and Evaluation
## Competence Assessment Battery

Paul D. Hood
Laird Blackwell
Barbara Havassy
Andrea Lash

FAR WEST LABORATORY
FOR EDUCATIONAL RESEARCH AND DEVELOPMENT

# MEMORANDUM

TO: Assessment Subjects

FROM: Laird Blackwell LRB and Paul Hood PH

DATE: May 2, 1975

SUBJECT: The Instructional Product Development and Evaluation Competence Assessment Battery

For the past year we have been developing a model assessment battery designed to assess both knowledge and skills requisite or useful in tasks that are frequently encountered in instructional product development. In several pilot tests, we have discovered that self-ratings, written examinations, and job sample simulations each have a degree of discriminant validity, but that each also contributes information not provided by the other.

We are now in the process of validating a revised version of the Competence Assessment Battery. Although the Battery focuses primarily on "entry-level" (M.A. degree level) competencies, in order to develop validation data, we have enlisted the cooperation of test subjects representing a very broad range of competencies, including students in graduate programs and highly-experienced R&D personnel located in R&D agencies throughout the country.

Normally, a student or on-the-job trainee would take the Battery after completing a related series of training modules which have been developed by the Far West Consortium for DD&E training. The information derived should provide the student and the instructor (or work supervisor in the case of an on-the-job training program) with some useful guidelines as to the degree of competence in several areas related to instructional product development. Please note that the current Battery makes no claim for comprehensiveness. For instance, a number of important skills and knowledge areas, e.g., task analysis, specifying learning sequences, are not included. Rather the Battery represents a model for developing test instrumentation that goes beyond the common written examination.

## Self-Rating

You are asked to complete two self-rating instruments. One consists of 61 statements of activities encountered in instructional research, development, dissemination and evaluation (RDD&E). This instrument has been designed to help us "profile" your own assessment of your "level of performance" in a broad range of RDD&E activities. A second, much shorter instrument, calls for you to rate your "level of knowledge" on 11 items which are most directly related to the knowledge and simulation tests of the Battery. (Please do not refer to your "level of performance" ratings while making these "level of knowledge" ratings; we have found in pilot testing that the two ratings

1855 FOLSOM STREET • SAN FRANCISCO, CALIFORNIA 94103 • (415) 565-3000

are highly, but by no means perfectly correlated.) Although the 61-item self-rating instrument includes activities over a broad range of R,D,D, and E, it emphasizes those involved in educational development and evaluation (some of which are directly related to those activities encountered in the Knowledge Test and/or simulation exercises). There is no time limit on the self-ratings. Most subjects will find that they require less than 1/2 hour to complete.

## Knowledge Test

Most of the questions in the Job Knowledge Test are designed to assess enabling knowledge for the simulation tasks. These questions should be useful in identifying gaps or weaknesses in your knowledge about educational development which would likely hinder your performance of some of the tasks a developer would frequently be required to perform. This test has a time limit of 1 hour.

## Simulation

In the series of four simulation exercises, you are asked to perform several tasks. All the necessary instructions and materials are included in written form. Time limits are suggested for each task; the entire series should take no longer than 3 1/2 - 4 hours.

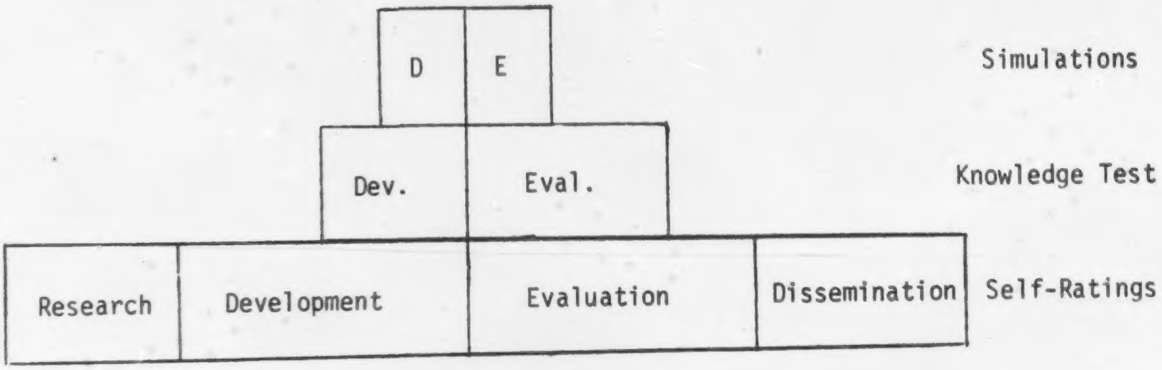
It can be seen that as the method of assessment approaches observation of on-the-job performance (i.e., as you go from self-ratings to knowledge test to simulation exercises), the scope of competencies assessed gets progressively narrower (see Figure 1 on the following page). If, as preliminary data analysis from field testing of the first form of these instruments suggests, each of these three instruments provides valid but largely independent information (i.e., each instrument discriminates between students and professionals, but has only modest correlations with the other instruments), the Competence Battery may be more useful as an example and guide to further instrument development than as an assessment tool for instructors and supervisors over a wide range of educational R,D,D, and E. Instructors/supervisors may want to construct knowledge items and/or simulation exercises relevant to their particular area of interest or concern.

In order to provide us with information on the validity and usefulness of this Battery, we would appreciate your filling out the DD&E Student Biographic Data Form regarding your relevant employment, education, training, etc., and then completing the self-ratings of performance, the self-ratings of knowledge, the knowledge test, and the simulations in this order. The entire sequence should take no longer than 5 - 6 hours; you may want to break this into two or three sessions.

Please note that all products of your work and all forms you complete will be treated as confidential information and will be used only for test development purposes. The set of response forms you have received are marked with an identification number so that the various pieces can be kept as one file. This is the only identification used. A supply of paper is provided for your convenience in completing the products called for in the simulation task. (You may use your own supply of paper if you wish.) Please be sure to mark your identification number on the top of each page of your products.

Figure 1

Coverage of R,D,D, & E Activities by Instruments in Competence Battery


D
E
Simulations
Dev.
Eval.
Knowledge Test
Research
Development
Evaluation
Dissemination
Self-Ratings

With your Competence Battery materials, you have received a contract to complete (so we can send you your payment) and a return envelope addressed to us at Far West Laboratory. To expedite our receipt of your completed Battery, please enclose all materials (i.e., completed biographic form, self-ratings, knowledge test, simulations, Task Evaluation forms, and signed contract) in the envelope and return it as soon as possible to us. Thank you for your cooperation.

LB/PH/cb
DD&E/CAB/4-75

## R,D,D,&E ACTIVITIES

These self-rating items focus on activities related to educational R,D,D,&E. For each item, indicate how well you feel you can perform the activity by writing the appropriate number from the following six-point scale on the line next to the item. (Please remove this page so you can refer to the performance levels listed below while you complete the ratings.)

### Level of Performance

1. I cannot perform this activity even with supervision or guidance.
2. I can perform this activity if I have either detailed instructions or close supervision.
3. I can perform this activity if I am given enough general supervision or general instructions.
4. I can perform this activity quite satisfactorily without supervision or job aids.
5. I can perform this activity quickly and efficiently and can do a top quality job.
6. I consider myself an expert in this activity and can accomplish unusually difficult or completely novel work.

DD&E/CAB/4-75

ID#______________

## R,D,D,&E Activities

Fill in the appropriate number from the scale on the previous page.

Items

__________ 1. Revise a test instrument on the basis of data collected during a pilot test of the instrument.

__________ 2. After product installation, interview users to determine if adequate information was provided about how to install and use the product.

__________ 3. Choose (or design) appropriate statistical techniques for data analysis.

__________ 4. Use bibliographic resources such as Research in Education and Current Index to Journals in Education to locate information supporting the need to develop an educational program or product.

__________ 5. Design a monitoring system that will provide data on the status of the operating system (such as actual vs. intended outcomes, unmet needs, problems, etc.).

__________ 6. Supervise professional R&D personnel.

__________ 7. From a general description of a product's purposes, specify appropriate terminal objectives.

__________ 8. Prepare a sample of an educational product which can be presented to a specified target group in order to determine if the proposed product meets the group's needs.

__________ 9. Review first draft materials to determine if they comply with specifications regarding objectives, sequencing, and content.

__________ 10. Derive a set of questions which could be sent with your product to reviewers in order to focus their review on critical aspects of your product.

__________ 11. Use and interpret statistical regression techniques.

DD&E/CAB/4-75

__________ 12. Identify the types of questions which can and cannot be answered by different research designs.

__________ 13. Design data collection procedures to maintain privacy or confidentiality in collecting, processing and storing information.

__________ 14. Conduct a search to determine if products exist which could meet your need(s) for instructional materials.

__________ 15. Confer with different types of users about their problems in the use of an educational product.

__________ 16. From a summary report of a tryout, specify the problems that need to be remedied by product revision.

__________ 17. Write and submit proposals to obtain funding and to negotiate with funding agencies.

__________ 18. Determine if new test instruments need to be developed by reviewing how well available tests fit the evaluation objectives.

__________ 19. Identify and articulate the problem in a research study.

__________ 20. Determine financial resources necessary to conduct a program or project and use accounting procedures to operate within a program or project budget.

__________ 21. Given the summary report of an initial tryout, suggest new methods to use in the second tryout to improve tryout validity.

__________ 22. Given a problem statement, information on the history of the problem, objectives and possible solutions, write a 20-page proposal for solving the problem, including a rationale for the approach and a development schedule.

__________ 23. Formulate testable hypotheses in a research study.

__________ 24. Prepare a coding scheme which will be used by a group of coders in preparing field test data for computer analysis.

__________ 25. Make arrangements to secure copyrights and copyright clearances where needed.

__________ 26. Given an educational product which you are developing, prepare tests and measures to assess student performance.

__________ 27. Identify factors which can jeopardize internal and external validity.

__________ 28. Convert a chapter in an instructional text into a branching programmed instruction sequence.

__________ 29. Determine the thoroughness of distribution which occurred in dissemination of an educational product.

__________ 30. In planning the evaluation of a new training program, determine the types of subjects which should be used.

__________ 31. Write a detailed product description including all the necessary elements.

__________ 32. Prepare a product design which is consistent with the stated overall purpose and use of the product.

__________ 33. Employ sampling theory and techniques to develop a stratified or clustered sampling plan.

__________ 34. With production personnel, determine the sequence of activities for production of instructional materials.

__________ 35. Review a draft field test report to determine if it is correct, complete and ready for release.

__________ 36. Deliver an oral presentation to a group of more than 20 professional educators to disseminate information about a project or product.

__________ 37. Plan interviews with potential users for the purpose of determining a market for your product.

__________ 38. Locate existing methods or strategies which can be used in potential product development.

________ 39. Write scripts for instructional films, videotapes, or sound filmstrips.

________ 40. Categorize instructional objectives in terms of a taxonomy (e.g., Bloom's Taxonomy of the cognitive domain or Gagne's conditions of learning).

________ 41. Collect information on installation costs for your instructional product.

________ 42. Determine an experimental design and statistical techniques whose underlying theoretical assumptions are consistent.

________ 43. Identify the crucial characteristics of a target group which may influence the dissemination effort.

________ 44. Provided with product specifications, the test product, field test reports, and expert reviews, make recommendations regarding the nature and extent of revisions required.

________ 45. Prepare a memorandum which summarizes the actions taken and problems encountered in a meeting with your development team.

________ 46. Write exercises which the learner should do in order to master concepts or principles in an instructional unit.

________ 47. Evaluate alternative media and instructional methods to match them with instructional objectives and learner characteristics.

________ 48. Evaluate the feasibility and risks associated with alternative approaches for redesigning a product that has failed to meet specifications.

________ 49. Design data collection forms for a marketing study of an educational product.

________ 50. Prepare a summary report of the test of your instructional product.

________ 51. In preparing a proposal for a development project, search for and select references which would support the rational and technical approach of the project.

__________ 52. Determine revisions that can be made within specified time constraints.

__________ 53. Review the relevant research literature for evidence to support the selection of an instructional strategy.

__________ 54. Determine the effectiveness of using a demonstration to disseminate information to your target audience.

__________ 55. Use management and planning systems such as PERT (Program Evaluation and Review Technique), PPBS (Program Planning Budgeting System), or Critical Path Analysis.

__________ 56. Evaluate at least three different market research techniques applicable to a specific development.

__________ 57. Specify a set of procedures to be followed by evaluation personnel in administering test instruments.

__________ 58. Analyze job or task requirements to determine objectives for an instructional program or product.

__________ 59. Determine questions you would like to have answered by the data collected in a tryout of your product.

__________ 60. Obtain information about strategies used to market educational products similar to yours.

__________ 61. Analyze discrepancies between expected and actual test outcomes.

ID#__________

## Job Knowledge Questions

Most of these questions were designed to assess knowledge that would be necessary or useful in performing some of the tasks of an educational developer. Use the space available to answer the questions briefly but completely, in many cases a simple list will suffice. If you have any problems or questions, ask the proctor or coordinator for help.

Please allow 1 to 1 1/4 hours to complete this task.

DD&E/CAB/11-74

1. What important information that should be included in a description of a product for a developer is missing from the following capsule version of such a description?

> The "Victorian England" course will serve as introduction to art, literature, and social life in England during the Victorian period. The course will include selected readings from texts on Victorian social life and the arts, but will emphasize the visual arts and literature of the period as instruction in those areas as well as sources of information about social customs. "Victorian England," when produced and packaged, will include the following: 1) a book of readings, for each student, including works by and about residents of Victorian England, and 2) a paperback book, for each student, of black-and-white photographs of Victorian painting, sculpture, and architecture.
>
> Users will be required to read the required text, but user activities will focus on analyzing and discussing reproductions of Victorian art and several of the best-known novels of Dickens.

2. On the basis of the following description, specify appropriate terminal objectives for this product.

> The "Plant Life of New Zealand" film will include still photographs as well as movie sequences of New Zealand's most common plants both wild and domestic. The film will be divided into several segments, each covering a different family of plants. Each segment will be followed by a short quiz or plant identification and associated knowledge. The film is intended to familiarize the high school biology or ecology student with New Zealand's common plants and their families and to teach the student what plants are edible and what plants are poisonous.

3. An instructional film on the behavior and attitudes a woman might adopt during pregnancy to help her prepare for child-birth has been developed by a community health service to be used in conjunction with reading, discussion, and practice in a "store-front" weekly course for pregnant women. In order to try out this film on as many people with as few administrative problems as possible, the developers plan to show it to several hygiene classes in a nearby high school and then test for knowledge as well as interview for attitudes toward the film.

   On the basis of this description, discuss briefly at least four aspects of this proposed try-out that could affect the relevance and importance of the information it would yield.

4. Why is a tryout useful before a field test? Why not omit the try-out and proceed to the field test as soon as possible?

5. Briefly compare the advantages and disadvantages of using photographs of real subjects vs. drawings ~~of~~ or cartoons for a film sequence.

(a)

6. A programmed text has been developed to teach high-school students the rudiments of statistical inference. Indicate three questions a developer might want to answer in a tryout about user satisfaction. Suggest two methods for collecting information on each of these questions.

7. What are the most important kinds of information a script writer needs to know in order to produce an instructional film?

8. Your R&D agency has been preparing a science unit for grades 2-4. The unit includes a movie, student workbook, and teacher guidebook, for each of six subjects presented. Listed below are questions which developers of the science unit would like to have answered by reviewers. For each question, suggest at least one type of person who could review the product to help answer the question (e.g., Media Specialist).

   1. Are the goals of the product valid and reasonable?

   2. Is the content provided correct?

   3. Is the media used appropriate for the target audience?

   4. Is the organization and sequencing of the content effective?

   5. Is the AV equipment required easily obtained by people who will be using the product?

   6. Is the language level appropriate for the target audience?

(a)

9. A large oil company has asked your R&D agency to develop a method of getting information about the uses and processing of petroleum and the present research taking place in the industry to upper elementary schools. A decision was made to develop a self-contained presentation in the form of a 16mm, 45 minute colored film. The film, shot on location, explains how petroleum is refined, why new methods of refinement are needed and what is being done to find new methods. In addition to the film, a short booklet for each student has been developed which shows, in a cartoon sequence, the many uses petroleum has in our society.

   Although you have been working with the content expert provided by the company, no one else outside your agency has seen your product. Below are three people who have agreed to review your product. For each reviewer, list at least three questions you would ask them about your product.

   1. A 5th grade teacher

   2. A sixth grade student

   3. The Oil Company Representative who initially asked you to produce the product.

10. If the subjects you have selected for the tryout of your product have difficulty in using the product effectively, what factors other than the product itself could have caused the difficulty. Specify three causes and, for each, suggest a remedy.

11. Outline the sequence of major steps (at least five but not more than ten) that are most important in the production of a sound filmstrip between (a) the definition of the instructional objectives and (b) the collection of the first validation (pilot test) data.

12. Intrinsic motivation (when the learner finds performance of the task itself rewarding) and extrinsic motivation (when a person's reasons for doing a task lie outside the task) can both be powerful forces influencing learning. Discuss how tests are used in instructional materials to enhance a) intrinsic motivation and b) extrinsic motivation.

(a)

Briefly define and describe the relationship to (or place in) educational development of each term listed below.

13. Field test

14. Enabling objectives

15. Competence domain

16. Content validity

17. Formative evaluation

18. Contingency management planning

19. Prototype

20. Constraints

(a)

# AN EXERCISE IN EDUCATIONAL DEVELOPMENT

Please read this before attempting Task 1.

DOAE/CAB/4-75

# AN EXERCISE IN EDUCATIONAL DEVELOPMENT

Competence-based educational programs need some means for assessing students' competency. Often the assessment depends primarily on tests of knowledge. However, it is desirable to assess performance as well. To some extent this can be done by evaluating the various products a student produces as the outcome of special projects or as part of their internship work. One problem in basing assessment solely on such products is that there are many uncontrolled factors; for instance, how much help was given or how much time was available. To overcome this lack of standardization, a simulation test has been developed.

Because this kind of simulation is an entirely new experience for many persons, we urge you to read this introduction. It may help to answer some of your questions and it should enable you to understand how the simulation is organized and what you will be expected to do.

This simulation will require you to perform many things that an educational product developer would need to accomplish. The tasks have been chosen to provide a "sample" of a number of specific development competencies that are frequently encountered in product development work. The tasks included in this simulation exercise have been designed to focus on essentials and bypass a lot of the less relevant detail. This has been achieved in several ways.

First, a relatively simple, and yet not trivial development task is used. You will find that the subject matter is sufficiently simple that no special content expertise is required. Second, the sequence of tasks is organized so that the information you receive in earlier tasks can be used in performing later tasks. This saves you from having to do a lot of reading about entirely unrelated tasks, but more importantly, you will find that the information available to you (although never enough to answer all your questions) will accumulate until, in the later tasks, you are confronted with a fairly realistic development situation which has a meaningful and relevant "history." Third, various kinds of information are provided (e.g., "models," instructions, lists of questions, etc.) which help to define what you are to do and which supply a continuity of sorts in terms of what "happened" between each task.

Something we've kept in mind in constructing this exercise is that development is a team effort, with usually some degree of specialization of work. Almost always you can count on being able to interact with others for advice or assistance. The information and instructions you will be given in this simulation attempt to serve as the interaction you would ordinarily have with your supervisor and other team members. Furthermore, you will find that several "team decisions" will be made during the simulation. These decisions in effect say, "Your work has been considered and the team has decided that this is what we should do next..."

You will find that nearly all instructions and "communications" have been preplanned and are in written form. Obviously, we are unable to anticipate every problem you may encounter. You should try to work on the basis of the written information and instructions given you. While your questions and comments are encouraged and even requested at certain points in this simulation exercise, we would like you to please try to confine your questions and comments to writing. But, if you have a significant problem that cannot be handled in writing, ask the test administrator for help.

Note that this is not a strictly speeded test, but there are time limits. At nearly every stage, you may feel that the time allotted is insufficient. Do the best that you can. Make sure you finish at least the essentials. The time limits have been set so that most beginning developers can complete at least a "rough draft" if they have been efficient in their use of time. Read your instructions carefully. Make sure you understand what you are to do. You may want to jot down how many minutes you can spend on each part of your plan of work, and then keep track of the time. Quality generally counts more than quantity, but if you fail to allow yourself the time needed to produce the tangible document(s) required at each stage of this simulation, you may find it difficult to complete the task within the time limit. If you finish early, review your work and make any revisions or additions which you think will improve it. Scoring will be based on the quality of your work accomplished within the time limit.

Please note that your performance will be assessed on several dimensions: completeness, appropriateness to instructions and background information provided, originality of ideas, sensitivity to problems and issues, technical quality of products, and quality of writing. Each task required of you focuses on a related set of product development knowledges and skills. The tasks cover a broad range of competence levels, but are focused primarily on the kinds of tasks a beginning developer (at least a college graduate with one year of experience or training in product development) would be expected to perform with some, but not a great amount of supervision. The total novice to product development may find the task difficult but not impossible.

We hope you find this exercise a challenging and useful experience.

* * *

You should now read "Your Role as a Member of the Development Team."

## YOUR ROLE AS A MEMBER OF THE DEVELOPMENT TEAM

Please imagine you have been employed by an organization which has subcontracted with the Far West Laboratory for Educational Research and Development to produce parent training materials for the Parent/Child Toy-Lending Library. A development team at the Far West Laboratory has been working to create the toys involved in the Toy-Lending Library. Most of the toys have already been tried out with children. The toys are either being revised or are ready for production. Your organization is developing a package of materials to be used to train parents of three- and four-year-old children how the toys and games can be used for learning episodes.

The team you have joined has been at work for a relatively short time. They have produced an initial version (a prototype) of the materials needed for the first parent training session and have tried it out with a few test subjects. The results have been encouraging.

You have been assigned to the parent-training session on the "Feely Bags" toy. In the course of your work you can count on the following kinds of assistance:

1. The team leader will supervise your work. However, because of other pressing commitments at the moment, you can expect only very general instructions and guidance.
2. The secretarial services pool will type your drafts of instructional materials, memos, letters, reports, etc. You must be sure to write legibly and to provide complete instructions.
3. Production Services will produce filmstrips from scripts; art work from sketches and do printing and packaging work.

4. A research and evaluation team has made arrangements to conduct all of the project field testing. You will need to communicate your initial (tryout) requirements to them.

In order to better acquaint you with the parent training sessions, your supervisor has referred to you the following section of the original proposal which describes the general form of the materials for the parent training sessions.

> *For each parent training session a separate self-contained package of materials, called a* ***Resource Kit,*** *will be produced. This Resource Kit will contain all the materials needed for one session to instruct parents in the use of one toy.*
>
> *The training provided by the materials in the kit should not take more than two hours. The materials should not use technical terms and should be written clearly (at no more than the 8th grade reading level). Any equipment (audio-visual or other) needed for the session should be the type available in most schools or community centers, or easily rented.*
>
> *The materials which will be found in each Resource Kit are:*
>
> 1. *A Librarian's (course leader's) Manual to provide all information necessary for a leader of one training session. Instructions to Librarians will be double-spaced, numbered, and listed in order as they are to be used. Any section of the manual designed to be read aloud will be triple spaced and boxed.*
> 2. *A filmstrip presentation showing a parent and child using the toy.*
> 3. *A toy and set of instructions for each parent.*
>
> *The Librarian will read the Librarian's Manual before conducting the session and will refer to it during the session. The parents will be required to listen to an introduction of the toy by the Librarian, observe the filmstrip presentation, and take part in a role-playing session led by the Librarian.*
>
> *After reading the Librarian's Manual, the Librarian should be able to plan the session, introduce the toy and its purpose, organize the role-playing session, review the Responsive Program principles, and answer questions asked by the parents.*

*After the training session, the parents should be able to describe in their own words:*

1. *The purpose of the specific toy.*
2. *The games which can be played with the toy.*
3. *At least one principle of the Responsive Program.*
4. *An example of one of the Responsive Program's principles found in the audio-visual presentation.*

The following section will provide background information you may need about the Parent/Child Toy-Lending Library.

# THE PARENT/CHILD TOY-LENDING LIBRARY

The Parent/Child Toy-Lending Library developed by the Far West Laboratory for Educational Research and Development is designed to help parents participate in their children's learning. The program shows parents of three- and four-year-old children how toys and games can be used for learning episodes and allows them to borrow toys as one borrows a book from a library. In essence, the developers intend to create a system to change parent's attitudes about their children's learning potential and about their own competency. In addition to developing certain toys, the developers envision a parent training program as an important part of the Toy-Lending Library and will offer an eight-week course for parents which will meet once a week for about two hours. Without this parent training, the product degenerates into a package of toys, none of which is unique in itself.

At each weekly meeting, a new toy will be introduced, the concept it attempts to teach will be explained, and parents will become familiar with ways of using the toy. The new toys are then taken home for a week. When the course ends, parents will be encouraged to use the library as a permanent source of additional toys, games, books, and records.

The program's goal is to help parents promote the intellectual development of children in a way that is likely to support the development of a healthy self-concept. It is designed to foster children's growth in concepts and skills such as color, shapes, problem-solving and verbal communication. The general goals of the training sessions include:

1. Parents will feel that they are more competent in helping their children learn some important skills and concepts.
2. Parents will feel that they can influence the decisions that affect the education of their children.
3. Parents will feel that the child is capable of learning and can be successful.
4. The child increases his competency as a result of the interaction with the parents.[1]

## Far West Laboratory's Responsive Program

The developers' goals for the Toy Library reflect the long-range objectives of the Far West Laboratory's Responsive Program, an educational system being developed for children aged three to nine. These objectives are "to help children develop a healthy self-concept as it relates to learning in the school and the home, and to develop their intellectual ability."[2] A diversity of educational experiences is sought by the Responsive Program, which rests on the assumptions that much learning takes place in the home, that formal education must be supplemented and that the educational program must be closely tied to the child's culture and background. The responsive environment leaves a great deal of autonomy to the child: adult-initiated talk should take second place to child-initiated talk, and the child can choose not to participate in group activities (as long as he or she does not disturb the group). In addition, the child's activities should not depend upon extrinsic rewards or punishments and they should help in developing a useful skill, concept or attitude.

---

[1] A Guide to Securing and Installing the Parent/Child Toy-Lending Library. Far West Laboratory for Educational Research and Development, Berkeley, California/1972, C5.

[2] A Guide to Securing and Installing the Parent/Child Toy-Lending Library. Far West Laboratory for Educational Research and Development, Berkeley, California/1972, B3.

The Toy Library was conceptualized as a way to help parents create a responsive environment for learning in the home. Its principles are those of the Responsive Program as a whole:

1. Free exploration: The child is free to explore any toy and to change the rules of any game he or she may be playing with the parent.

2. Self-pacing: The child is free to work at his chosen speed and to stop work when the game or toy is no longer of interest.

3. Self-correction: The toys are so constructed that the child can immediately find out the results of explorations, either from the toy itself or from the parent.

4. Discovery learning: The child is given time to discover things for him or herself--the parent helps the child to think through a problem rather than giving the correct answer.

5. Self-reward: The learning activities are satisfying to the child. The game is not played because the parent will reward or punish for not playing, the child learns because he wants to.

The principal developer of the Responsive Program sees the Toy Library as a way of enhancing parents' involvement with their children's education. One staff member states that the developer:

> ...was probably one of the first to come out with the notion that parents are teachers too, and that parents could do some things in their homes that were being done in the Headstart, daycare, and private nursery schools. The program was to pull together some positive interaction between parent and child. This is not to say that parents are not positively interacting with their children already, but that they could be trained to do some skill development and concept development, and in so doing, know more about the child's educational ability and the way he learns and his whole mode of interacting.

As parents learn to be more active participants in their children's learning, it is thought that they will assert themselves more strongly in school decision-making. The developer points out:

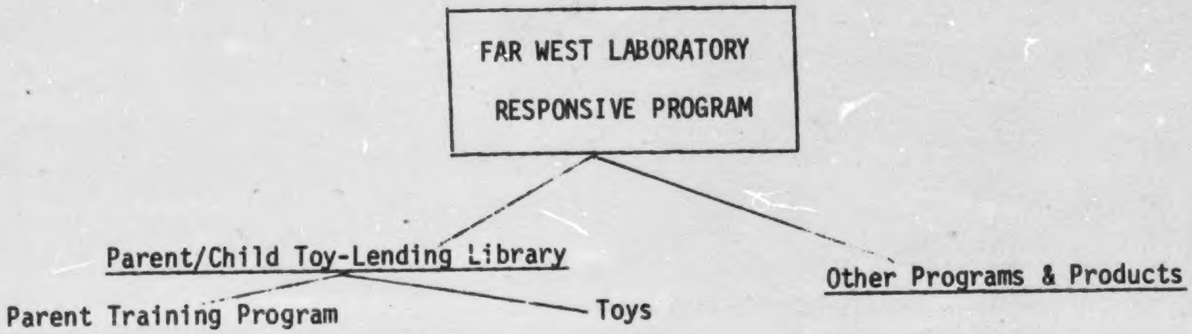
> "We hope that one of the functions that the training will serve will be to give the parents more control over the educational system--how to make it more responsive to their needs..."

The program is designed to fill a need for families above the income level required for Headstart programs but not affluent enough to afford nursery school tuition. The developers estimated that three-fifths of the parents in the country fit this description. Because of the training component, however, one family will not be able to act alone to purchase "the product." However, any one of a variety of organizations could set up a Toy Library. The Toy Library Program requires a course leader, a room for the weekly sessions, some audio-visual equipment and storage space. Schools, daycare centers, public libraries, churches, industries and unions are among the target buyers. If no already-organized group in a community wants to take on the program, a group of parents could incorporate as a nonprofit association and begin to operate a Toy Library.

The Toy Library program is meant to supplement the customary procedures for early childhood education. It establishes new roles for parents--the users, for organizations--the purchasers, and for the course leader--the facilitators. Furthermore, the developers contend that the physical products--the toys and other materials--are meaningless without the training program, which requires planning and money to set up.

## Summary

The Toy-Lending Library is a product which is part of the Far West Laboratory's Responsive Program. The Toy-Lending Library itself has two sub-products or components. These are (1) the toys themselves and (2) the parent training program. The relationship of these programs and products may be seen as:


FAR WEST LABORATORY
RESPONSIVE PROGRAM
Parent/Child Toy-Lending Library
Other Programs & Products
Parent Training Program
Toys

You have now completed the introduction material for this simulation. Turn to Task 1, Preparing Guidelines for Developers of the Librarian's Manual.

## TASK 1: PREPARING GUIDELINES FOR DEVELOPERS OF THE LIBRARIAN'S MANUAL

Time allowed (including reading) - 1 1/4 hours

# MEMORANDUM

**TO:** Members of Team Assigned to Prepare Guidelines for Developers of Librarian's Manual

**FROM:** Doris Shin, Team Leader

**SUBJECT:** Preparation of Guidelines

In the weeks ahead, our Parent Training Materials team is responsible for producing eight Resource Kits, one for each session of the Parent Training Program. Early in the Toy Lending Library project, a sample Resource Kit was developed for the first session of the Parent Training Program. Although the results from the testing of this first Resource Kit suggest a need for certain revisions, I think its format would be a good one for us to follow in developing the rest of the Kits.

For the Resource Kit on the Feely Bag toy, we need to prepare the Librarian's Manual for the parent training session, a filmstrip presentation, and instruction sheets for the parents in how to use the toy. The task of preparing the parent instruction sheets has been assigned to George Gregory, so the rest of us need to work on the Librarian's Manual.

I would like you to prepare a memo to the team who will actually develop Section 2 of the Librarian's Manual. Section 1, which is attached for you to read, gives general background information. In contrast, Section 2, which you should describe in your memo, will give step-by-step instructions that the Librarian can follow in teaching parents to use the toy. In order to develop this section, the team will need your memo to provide: 1) descriptions and specifications the developer can follow regarding purpose, format, objectives, user (i.e., Librarian) activities, constraints, and other relevant information, and 2) a brief outline of the main steps to be followed by the Librarian during the instructional session.

By referring to the attached documents and those you have previously received, you should have all the information you will need to prepare this memo.

### *TASK 1: PREPARING GUIDELINES FOR DEVELOPERS OF THE LIBRARIAN'S MANUAL*

*This task calls for the preparation of a memo providing guidelines for the developers of Section 2 of the Librarian's Manual for the parent training session on the Feely Bag toy. You will have 1 1/4 hours to complete it. It should be no longer than the equivalent of three (3) double-spaced typed pages (900 words). You have been provided with several documents (e.g., Feely Bag toy description and Section 1 of the Librarian's Manual--both attached to this memo, a description of the Toy Lending Library, a description of your role as a member of the development team) which should provide you with the information you will need to accomplish this task. You are free to re-read, refer to, and use these documents in any way during this part and later parts of the simulation. However, please do not refer to documents included in later tasks.*

*Remember, you are not responsible for actually developing the Librarian's Manual. Rather, you need to extract the relevant information from the documents provided and organize it in a memo so that the development team will know for whom the manual is to be written, in what form it is to be written, what it is supposed to accomplish, etc. Your memo should also outline the content of Section 2 of the manual, i.e., the steps the Librarian should follow in the parent training session (e.g., introducing parents, leading role-playing, etc.)*

REMINDER. A description of the Feely Bag toy and the background information section (Section 1) of the Librarian's Manual are attached. When you have finished Task 1, please fill out a copy of the DD&E Simulation Task Evaluation Form (goldenrod paper) before turning to Task 2.

## FEELY BAG TOY

### EQUIPMENT:

Small drawstring bag and two sets of cut-out shapes. Each set has these four shapes: a circle, a square, a triangle, and a rectangle.

### PURPOSE:

To help the child recognize shapes by touch and by sight.

### RESPONSIVE PROGRAM PRINCIPLE TO BE DISCUSSED IN CONJUNCTION WITH THIS TOY:

Self-correction - The child can immediately find out the results of explorations, either from the toy itself or from the parents.

### GAME TO BE PLAYED WITH THIS TOY:

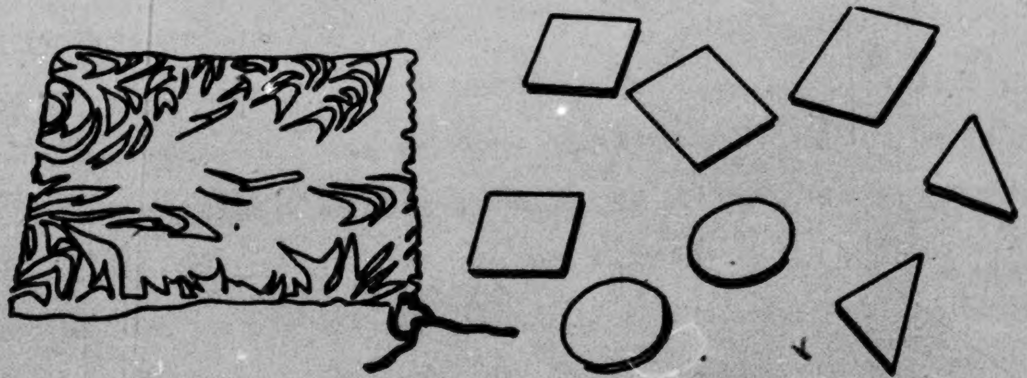
One set of shapes are placed in the drawstring bag. Parents are instructed to show the child a circle, square, triangle or rectangle from the second set of shapes, tell the child the name of the shape, and ask the child to find a shape in the bag that is the same as the shape the parent has.

### GENERAL INSTRUCTIONS FOR PARENTS:

A. Ask your child only once each day if he (or she) wishes to play the game.

B. The child may change the rules of the game at any time. You must follow the child's rules if he/she changes them.

C. You should stop the game when the child seems to lose interest.



3

# LIBRARIAN MANUAL

## SECTION I: HOW TO OPERATE A PARENT/CHILD TRAINING COURSE

# Section I

## HOW TO OPERATE A PARENT/CHILD TRAINING COURSE

A great deal of planning has to be done before a course for parents can be started. If all these plans are carefully followed, it will be possible to train many groups of parents to use the toys at home with their 3-year-old or 4-year-old children. This part of the manual includes all the information that is needed before the course begins.

### 1. What Does the Course Try to Do?

The Parent/Child Training Course teaches parents how to use toys and games that help preschool children learn some very important skills and ideas. In the course, the parents also learn some basic things about how children learn to think.

Children must develop the skills of seeing, hearing, feeling, and so on--because these are the things that help them think clearly and well. They must develop their language skills because language is the basic tool for thinking. They must develop their ability to form "concepts" (bigger ideas) because concepts help to organize thinking. They must develop problem-solving ability because problem-solving is the key purpose of thinking.

In the course, the parents also learn some ways of understanding how children feel about themselves, about their families, and about the world. It is very important for children to:

- feel good about themselves, their family, and their people;
- believe that what they think and say and do makes a difference;
- believe that they can be successful; and
- believe that they can solve problems.

As the parents begin to play these games with the child, they will begin to take a more active part in the child's education. At the same time, the parents will begin to make important decisions about what the child is going to learn and who is going to teach the child.

### 2. Who Developed the Games--And Why?

For several years a group of men and women at the Laboratory for Educational Research and Development have been working on a program called "The Responsive Program." The Laboratory is a non-profit public institution that operates mainly with money from the United States Department of Health, Education and Welfare. But the money to work on this program came from the Carnegie Corporation. The people who developed these games have worked for many years with parents and children and teachers. All of the toys used in the course have been tested and checked and tested again with many groups of parents and children in many different places.

The people who developed the games believe that:

- children should not get any special rewards for learning something new--the only reward should be the pleasure the children get from the act of learning;
- children should be free to "explore" what they are going to learn;
- children should learn at their own speed;
- children should learn right away what happens if they act in a certain way; and
- things that children play with should lead them to discover other things in the world around them.

All the toys used in the Parent/Child Training Course were tested very carefully to be sure that they met these basic needs.

3. What Materials are Needed for the Course?

There are eight Resource Kits available for the Parent Training course, one for each weekly meeting. The Resource Kits contain all materials (except audio-visual equipment) which are needed to train parents in the use of one toy. Each kit contains:

1. A Librarian's Manual for the course leader.
2. A filmstrip presentation.
3. A toy and set of instructions for each parent.

When ordering the Resource Kits, be sure to specify the number of parents who will be in the course, so enough toys and instructions are included in the Kit.

Be sure to collect the toys and instructions from the previous week at each session so they can be returned to the Resource Kit and used again. There should be no cost to the parent if the toy is lost or broken, but be sure the parents understand that the toys should be returned so others can use them.

4. What Other Equipment Is Needed?

The Parent/Child Course uses eight filmstrips (there is one for every toy) that show parents how to use each toy in several ways. With each filmstrip a sound tape is used so that the filmstrips will seem to be "how-to-do-it" movies.

To show the filmstrips, the course leader needs:

- A filmstrip projector.

  One can be borrowed from a nearby library or college or school, or church, or one can be rented from a local dealer (See "Audio-Visual Equipment and Supplies" listing in the yellow pages of the local phone directory).

- A cassette tape recorder.

  One can be borrowed from a friend, or from a nearby library or college or school.

The course leader must learn how to operate these machines before the first session of the course begins. The leader must check electrical wall sockets. Also, the leader must be sure to have a large piece of white paper taped on the wall or obtain a screen so the pictures will be easily visible. The filmstrip projector and the cassette player are run at the same time. One gives the pictures. The other gives the sounds.

### 5. Where Can the Course Be Located?

The meetings of the Parent/Child Training Course should be held in the same place each week. A room may be found in any of these locations.

- School Building
- Church
- Library
- Hospital
- Vacant Store
- Community Center
- Town Hall

It is very important that the location be within easy walking distance for the parents. A place near a local bus stop would be good, too.

### 6. How is the Course Leader Chosen?

The parents of the community will probably want to learn from someone who lives in that same community. They are not likely to be comfortable with an "outsider." The course leader should be someone who is really a part of the community. The amount of money the course leader earns should not be very different from the amount earned by the parents who take the course. But it is not necessary to be a mother or father to be a course leader. A warm, friendly man or woman--even one who is not a parent--can probably do a good job.

## 7. How Are Parents Selected for the Course?

If the course leader lives in the community, it won't be hard to find parents and ask them to come to the weekly meetings. Names of parents with very young children can be collected by talking to the office workers at the local elementary schools and high schools. The local librarian may have some names to suggest. If other community programs are already going on, the aides or parent coordinators may have good ideas. But, because schools already have to count ahead of time all the children who will be coming into their kindergartens, the school office is the best place to get the biggest list of names of parents of preschool children.

Parents cannot be signed up for the course by writing letters to them. Each parent must be told about the course personally. Getting a group of parents who are willing to take the course the first time is the very hardest job the leader will have. For that reason the leader should run only one course at the beginning. Later on, with more experience, two or three courses could probably be managed during the same period of time.

Child-care during the course meetings will always be a big problem. The child should not come to the meetings with the parent. Possibly a teen-ager can be given the job of taking care of some of the children so that the parents will have free time to learn.

A local school district or community college may be willing to give the parents adult-education credit for taking the course. The course leader should visit the office of the nearest adult-education program and show them this book. When the staff understands what the course is trying to do for parents and children, it will be able to decide quickly if credits can be given. And the chance to earn adult-education credits should make more parents want to come to all the meetings once they begin.

## 8. How Can Visitors be Handled?

No visitors should be allowed before the third or fourth meeting. It takes time for the parents to begin to feel comfortable with each other and with the course leader. But once the meetings begin to run smoothly, visitors who drop in should be made to feel welcome.

If, by chance, a parent brings a child to one of the sessions, the course leader may be able to use that child as a "model" in the training.

## 9. How do You Prepare for a Weekly Meeting?

Here is a plan for the 8-week course.

During the week before the session, read the Librarian's Manual for that session. The Librarian's Manual will explain what to do at each session and will offer suggestions of what to say to the parents.

Before each session begins, be sure to set up and check the filmstrip projector and the cassette tape player. Be sure the tape is in the position where the sound track starts. Focus the filmstrip projector on a blank section of the wall or on a large piece of white paper taped or tacked on the wall.

And, of course, be sure enough toys are ready so that each parent will be using the same toy at home during the following week.

Finally, be sure there are enough printed instructions for each game so that each parent can have everything needed to succeed.

## 10. What Happens in a Weekly Meeting?

As the course leader you are responsible for setting up the activities of each session. The Librarian's Manual for the session should help you do this. Each week you will introduce a new toy and describe its purpose. You can then show a filmstrip presentation of a parent and child using the toy. After the presentation, a kind of "acting" called "role-playing" is used: (a) to help each parent get a better idea of how each game works and (b) to help each parent begin to see what the child will learn from that special game.

Role-playing means pretending to be someone else. It means getting the "feeling" of another person's part in the game. One person takes the other person's place and acts as he would probably act. Playing a "role" lets the "actor" find out other people's ideas and feelings. (Children are "role-playing" when they dress up in adult clothes and pretend to be their parents.)

In the course, the parent is really doing the same kind of thing--except now the adult is pretending to be a child. In the course, role-playing lets the parent imagine how a child will act and feel in the game they will soon play together at home.

As the course leader you will also need to answer questions which the parents may have. The Librarian's Manual should help you in answering these questions. At each session you may wish to schedule time where the parents can discuss the toys and their experiences with them.

## TASK 2: PREPARING FOR THE TRYOUT OF THE RESOURCE KIT

Please do not read this task until you have completed Task 1 and an evaluation form for Task 1.

Time allowed (including reading) - 1/2 hour

# MEMORANDUM

**TO:** Designers of Librarian's Manual for Feely Bag Resource Kit

**FROM:** Doris Shin, Team Leader

**SUBJECT:** Tryout of Feely Bag Resource Kit

On the basis of the design for the Librarian's Manual which I put together to incorporate the best features of all your designs, a prototype of that manual has been prepared by other members of our development team. (I have attached the design and the prototype for you to look at.) It is now time for the Resource Kit on the Feely Bag toy (including the Librarian's Manual) to be tried out by potential users. Our Research and Evaluation Team will conduct the tryout. They intend to tryout this Kit at at least two different sites with a different Librarian for each site. They are going to have each Librarian attempt to recruit about six parents for the session.

Even though the Research and Evaluation people are going to assume main responsibility for the tryout, we need to provide them with some information. First, given the small number of people with whom the materials will be tried out and our restricted travel funds, please describe briefly what you think the Research and Evaluation people should look for in establishing test sites, i.e., what kind of sites, Librarians and parents. A paragraph should be sufficient for this.

Second, the Research and Evaluation team need to know what kind of questions we need to have answered in the tryout. I have been working on the questions for the filmstrip presentation and the session as a whole; I would like you to list the questions you think are important to have answered about the Librarian's Manual. List separately questions that you would ask: (a) the Librarian, (b) the parents, and (c) an observer of the session.

## *TASK 2: TRYOUT OF LIBRARIAN'S MANUAL*

*In this task you are to: (1) describe what the Research and Evaluation Team should look for in selecting test sites, librarians, and parents and (2) list the questions (separately for librarians, parents, and observers) that you would like to have answered in a tryout of the Librarian's Manual that we have developed.*

*The list of questions is the more important of the two activities. It should take no longer than the equivalent of two (2) typewritten double-spaced pages.*

*The description for the Research and Evaluation Team should be brief--one or two paragraphs.*

*This task should take no more than 30 minutes.*

REMINDER. The design and prototype of the Librarian's Manual are attached. When you have finished Task 2 please fill out a copy of the DD&E Simulation Task Evaluation Form (goldenrod paper) before turning to Task 3.

# DESIGN OF LIBRARIAN'S MANUAL FOR FEELY BAG TOY SESSION

The purpose of this manual is to provide instructions to the librarian for leading the session of the Parent Training Program which introduces the toy called FEELY BAG. This toy is designed to help the child recognize shapes by sight and touch.

After reading the manual, the librarian should be able to: introduce the toy FEELY BAG, run the filmstrip presentation, discuss the Responsive Program's Principle which is illustrated in the filmstrip, lead the role-playing session, and answer questions the parents may have.

The manual should be written at the reading level used in the first librarian's manual. The format should be similar to the format of section two of the first manual. The instructions should be double-spaced and numbered. Any examples which could be read aloud should appear beneath the instructions, be triple-spaced and boxed.

The librarian will need to refer to the manual during the session in order to follow the instructions and to run the filmstrip presentation. To help the librarian run the filmstrip presentation, a copy of the script indicating when frames are to be changed must be included in the manual.

The instructions should provide information on:

1. Beginning the session by reintroducing the parents.
2. Leading a discussion about the parents experiences with the toy from Session 1 (SOUND CANS).
3. Collecting toys and parent instructions from Session 1 (SOUND CANS).
4. Introducing FEELY BAG and its purpose.
5. Showing the filmstrip presentation.
6. Discussing important points illustrated in the presentation, emphasizing the Responsive Program Principle-Self Correction. (The child can immediately find out the results of explorations, either from the toy itself or from the parents.)

7. Distributing the toy FEELY BAG and parent instructions.
8. Leading the role-playing session.
9. Reviewing the purpose of the FEELY BAG toy.
10. Ending the session with a question and answer period.

The manual should not be more than ten pages. It should be three-hole punched so that it can be placed in a small folder.

# LIBRARIAN MANUAL

## SECTION II: SESSION II - FEELY BAG

DD&E/CAB/4-75

## Section II

## SESSION II - FEELY BAG

Instructions for the first training session are listed below. Examples of what to say to the group of parents are provided in the boxes. You may read these examples aloud, or use them to plan what you will say to the group.

1. Introduce yourself as the course leader.

> For those of you who may not already know me, my name is ____________________. I will be meeting with you for the next several weeks to help you learn how to use the toys we have here with your children.

2. Ask the parents to introduce themselves.

> Since we'll be working together in learning about these toys and and the ways children learn, it is important that we get to know each other. Why don't we start by having each of you introduce yourselves to the group? Let's begin (to my left, with the first row, etc.)

3. Tell where the Toy Lending Library originated, and what it's purpose is.

The Toy Lending Library was developed at the Far West Laboratory for Educational Research by people who specialize in teaching young children. These people felt that children can and do learn in many places other than school. They felt it is important for parents to take part in their child's education, and for children to have fun while learning things they will need to know. Each of the toys in this program was designed to teach an important skill such as hearing, feeling, counting or problem solving. All of the toys have been tried out several times with many groups of parents and children in many different places. In the next several weeks you will be learning how to use these toys to help your children learn new skills and enjoy themselves while they are learning.

4. Describe what the weekly meeting will be like and what the parents will be expected to do.

At the weekly meetings I will distribute a new toy to each of you. We will see a filmstrip showing a parent and a child using the toy, and then discuss how to use the toy with your children. We will also have a chance to tryout the toys ourselves. After the meeting you can take the toy home and let your children play with it. The toy must be returned the following week so it can be used again with other parents. Hopefully we will have time to talk about the experience you and your children had with the toy when you return it. After the eight meetings you will be able to borrow other toys as you would borrow a book from the library. Each of the toys will have instructions with them so you will know how to play games with your children when using the toy.

5. Introduce the toy FEELY BAG and describe its purpose.

This toy is called FEELY BAG. It is made up of a small drawstring bag and two sets of cut-out shapes. Each set has these four shapes: a circle, a square, a triangle, and a rectangle. [SHOW THE PARENTS THE SHAPES AS YOU MENTION THEM] This toy was designed to teach children to recognize shapes by touch and by sight. FEELY BAG should help your children learn to tell when shapes are the same and when they are different by looking at them or by feeling them. An important aspect of this toy is that it is self-corrective; the child can find out the results of his/her explorations immediately either from the toy itself or from the parents.

6. Show the filmstrip.

> Now I will start the filmstrip which shows how a parent uses FEELY BAG to play games with a child.

7. Discuss important points illustrated in the filmstrip. (How to introduce a toy to a child, what to do if a child makes up rules, what to do if a child picks out the wrong shape.)

> I think the filmstrip showed some important things to remember in playing games with children. If you recall, the parent did not begin by telling the child about the game they would play. Instead, she let the child play with the shapes by herself. Even after the parent tried to start the game she did not force the child to play. The little girl wanted to play her own game and the parent let her do just that. It is important for children to explore what they are going to learn and to enjoy learning. This is why the parent in the filmstrip allowed the child to play with the toys by herself and did not play the game until the child wanted to.

Another point the filmstrip illustrates is the use of specific words when describing things to your child. When the girl in the picture chose the wrong shape, the parent said, "These two shapes are not the same, try again" instead of saying, "That's the wrong shape." When the child did choose the right shape the parent said, "Yes, these two shapes are the same; they are both circles." In both cases the parent told the child why the shape was right or wrong by using specific words.

You will find some of these hints to use when playing with your child in the instructions which I will pass out with the toys.

**NOTE:**

If the parents have not started asking questions or discussing the filmstrip, mention that they are welcome to make comments at anytime by saying something like:

It seems I've been doing most of the talking so far. I hope we can keep these meetings rather informal, so I want you to feel free to ask questions or make comments at anytime. Does anyone have a question or something they would like to discuss with the group?

8. If the parent do not have any questions, or if the group is finished discussing their questions and concerns, give each parent a set of instructions and a toy.

9. Introduce the Role-Playing Session.

> Before you play the game with your children, it would be good to practice with it. We'll divide the group into pairs and practice with each other. This practice is called role-playing. Role-playing means that a person is acting the part of another person. The actor is doing what another person would do in a real situation. Acting the part of another person helps you to understand how other people do things and what it feels like to be in "someone elses shoes." Each of us will get a chance to act or role-play the part of a child and the part of a parent. Acting the part of the parent shouldn't be very hard, but acting the part of a child may be hard at first. Try to pretend you are three years old and your parent is showing you a new game.

NOTE:

Before dividing the group into pairs, ask one parent to role-play with you. Take turns being the child. After the others have seen what role-playing is like, ask if there are any questions, and then divide the group into pairs. Remind them to let the child explore and to use specific words when talking to the child.

10. After each parent has had a chance to play the role of a child, bring them back together as a group, ask if they have any questions, and review the game FEELY BAG.

> O.K., Now that we've all had a chance to feel what it is like being the child, It's get back together as a group. Did you run across any problems trying to use specific words with the child, or letting the child explore the toy?

(If the parents feel like discussing their role-playing experience, allow them time to do this. After they have finished, review the purpose of FEELY BAG.)

> I'd like to review FEELY BAG now, before we leave. It's a toy which is designed to teach children how to tell when shapes are the same and when they are different by sight and by touch. It should be fun for your children to play with the toy. Allow them to explore the toy, and to make-up rules and games. Remember to use specific words when your're talking to your children. If your child gets tired or doesn't want to play anymore, then set the game aside and ask the child

to play another day. Don't force the child to play, because we want them to have fun when learning.

11. End the session by asking if there are any questions. Check to see everyone has a toy and instructions. Remind the parents to read the instructions before playing the game, and to bring both the instructions and toy back next week.

## FEELY BAG SCRIPT

| NAME NUMBER | AUDIO |
|---|---|
| 1 | This is the first game with the Feely Bag.* |
| 2 | To play the game, you'll need the Feely Bag and two circles, two squares, two triangles, and two rectangles.* |
| 3 | The Feely Bag games help children recognize shapes by both touch and sight.* |
| 4 | Begin by putting out the Feely Bag and shapes.* |
| 5 | Let your child play with them for a while.<br>(Child) "This a circle? I can't. Where is it?"* |
| 6 | To start the game, put one of each shape in the bag.* |
| 7 | Put the other four shapes on a table or on the floor so your child can see them.* |
| 8 | Pick up any one of the shapes.* |
| 9 | For example, you might pick up a circle and say,* |
| 10 | "Find a shape in the bag that is the same as this circle."* |
| 11 | If she chooses a different shape, hold up your circle next to the one your child took out of the bag.* |
| 12 | Tell her, "These two shapes are not the same; try again." And put aside the shape she took out.* |
| 13 | (Child) "I try again--I'll find one yet--I'll find one yet--me find one."* |
| 14 | (Child) "Uh--I found one already--"* |
| 15 | Since she did choose a circle, tell her, "Yes, these two shapes are the same; they are both circles."* |
| 16 | (Silence)* |
| 17 | Since she found the correct shape, put all four shapes back into the Feely Bag again.* |
| 18 | Now pick up another shape--for example, a triangle.* |

---

*These marks show when the "tone" or "beep" is heard. Advance to the next frame each time at the tone signal.

| FRAME NUMBER | AUDIO |
|---|---|
| 19 | Say to her, "Find the shape in the bag that is the same as this traiangle."* |
| 20 | If she peeks into the bag during the game, tell her, "Now find the shape without looking.* |
| 21 | If she pulls out a triangle...(Child) "I found it." "These two shapes are the same; they are both triangles."* |
| 22 | Put all four shapes back into the bag...then continue to play the game.* |
| 23 | "This is a square. Find a shape that's in the bag that is the same as this square."* |
| 24 | (Child) "I can't, I don't know where to find it." "These two shapes are not the same, try again." (Child) "I'm trying to find it; here it is. I don't know---"* |
| 25 | "These two shapes are the same; they are both squares."* |
| 26 | If your child thinks of new rules for this game, let her play it her way.* |
| 27 | Here she's invented her own special way to use the Feely Bag.* |
| 28 | Notice how the parent goes right along with her, changing from one game to another.* |
| 29 | (Child) "What's this thing doing in this for?"* |
| 30 | Stop playing as soon as she seems to be bored or tired.* |

-END-

## TASK 3: RECOMMENDING REVISIONS FOR THE LIBRARIAN'S MANUAL

Please do not read this task until you have completed Task 2 and an evaluation form for Task 2.

Time allowed (including reading) - 1 hour

# MEMORANDUM

**TO:** Developers of the Feely Bag Librarian's Manual

**FROM:** Doris Shin, Team Leader

**SUBJECT:** Revising the Librarian's Manual

I have just received the report on the tryout of the Feely Bag Resource Kit. You will find a copy of the report attached to this memo.

There are two things I would like you to do with the information in this report. The first and most important is to list your suggestions for revising the Librarian's Manual based on the results of this tryout.

The second is to critique the tryout by answering the following questions:

1. What are the major strengths of the design and implementation of the tryout?
2. What are its major weaknesses?
3. If the Research and Evaluation Team were to undertake a second tryout of this resource kit, what additional instructions would you given them regarding:
    - °selection of test site(s), librarian(s), parents?
    - °making arrangements for the tryout?
    - °directions which should be given to the librarian(s) before the session?
    - °specific directions and questions for the evaluator observing the session?

## *TASK 3: RECOMMENDING REVISIONS FOR THE LIBRARIAN'S MANUAL*

*There are two activities involved in this task. Listed in order of importance they are:*

1. *Recommending revisions for the Librarian's Manual. (This product should be no longer than the equivalent of three (3) double-spaced typed pages.)*
2. *Critiquing the tryout. (This product should be no longer than the equivalent of two (2) double-spaced typed pages.)*

*You have 1 hour to complete this task.*

REMINDER. The Summary Report of the tryout which you will need to accomplish Task 3 is attached. When you have finished Task 3 please fill out a copy of the DD&E Simulation Evaluation Form before turning to Task 4.

## SUMMARY OF THE TRYOUT OF THE FEELY BAG RESOURCE KIT

This unit was pilot tested at the same two sites used to test the prototype lesson: the Neighborhood Care Center in the San Francisco Mission District and at the West Oakland Nursery School. A member of the staff was recruited at each organization to serve as the toy librarian. This person agreed to recruit six parents to attend a toy training session.

Neighborhood Care Center. This Center is located in the Mission District, an inner-city, multi-ethnic neighborhood.* The Neighborhood Care Center has been in operation for three years, with funds supplied by United Crusade, federal, and city sources.

Mrs. Lopez, who is one of three part-time paid staff members (paraprofessionals) and has been with the Center for two years, was recruited to serve as the librarian. She was familiar with the project since she had served as a "parent" for the prototype test. The materials (librarian's guide, AV equipment, filmstrip, tapes, toys, etc.) were delivered the day before the scheduled session. We helped Mrs. Lopez check the material and determined that she knew how to operate the AV equipment.

Mrs. Lopez reports that she spent two hours reading and preparing for the training session, which was scheduled for the following evening at 7:00 p.m. I arrived at 6:30. Mrs. Lopez had already set up the equipment in one of the playrooms. By 7:15 five parents arrived. (The sixth, had called to say she couldn't come). Two were Latino, two were Black and one was Anglo. All had children enrolled in the school. They obviously knew Mrs. Lopez and each other. The session lasted from 7:20 to 8:30. By arrangement, Mrs. Lopez served refreshments while I held a brief interview with each parent individually. After the parents left, I interviewed Mrs. Lopez.

---

* The Census Bureau statistics show the 1970 population of the area as 32% Anglo, 48% Spanish surname, 5% Black, 2% Indian, 10% Other Specified (generally, persons from the Pacific including Samoa, Guam and the Philippines), and 3% Other.

Results. Mrs. Lopez followed the recommended sequence to the last detail, except that she had to explain what the Toy Library was, that this was a try-out and why I was there. This would normally have been handled in the first session. We should have anticipated this problem. Suggest we produce a very short orientation for the librarian to use whenever any of the manuals are tried out with a new group of parents. This wouldn't be encountered in a regular series of sessions but was a problem at this site. Fortunately, Mrs. Lopez handled the situation adequately.

The filmstrip presentation got good attention and seemed to serve as a useful orientation. Getting parents to handle toys themselves worked out well, But the role-playing was a disaster on the first two trials. Neither of the parents could get the idea of playing a child's role. Finally, it worked when Mrs. Lopez played the child's role with one of the participants playing the parent. I'm not sure whether there is a language problem, a cultural problem, or whether they needed a better model. (Say, a filmstrip depicting role-playing.) Best guess is its language and model. Both of the parents who played the child's role said they didn't know what was expected of them.

The interview results indicate that all five parents were favorable to some degree. All of them could explain the purpose of the game and each wanted to try it with their children. However, only two of the five parents could describe the Responsive Program Principle covered in the lesson in their own words, even after I gave one example.

Mrs. Lopez said she thought the instructions were easy to read and easy to understand. She thought the filmstrip could have lasted longer. She said that she had never heard of role-playing and that the description in the manual was too short to give her a very good idea of what to do. She would also like to see a "question and answer" section so she would have better answers for some of the parents questions.

This seems to be a good idea. Questions asked at this session:

- Are there other games to play with the toy?
- What do I do if a piece is lost or broken?
- How long should I play the game?

West Oakland Nursery School. The school is a cooperative located in a low-income area in Oakland* which has operated for five years. Mrs. Adler, who has two children at the center, was recruited. She was interested in the idea but unsure of her ability to act as the librarian. The advance preparation was similar to that reported above for the Neighborhood Center. Only four parents appeared for the 10:00 a.m. session. Two white and two Black. Because of our experience at the Neighborhood Center and at Mrs. Adler's request, I spent about 10 minutes explaining the toy library project and the purpose of the tryout. (Recognize this was not what was planned, but it seemed necessary to set the stage and put Mrs. Adler at ease.)

Mrs. Adler started by giving each parent a set of the toys and then she talked about them for nearly fifteen minutes. The parents got restless. Finally, when the filmstrip was shown, interest picked up. Had Mrs. Adler followed the suggested schedule, I think this could have been avoided. Role-playing session again got off to a slow start. Couldn't get first two parents to play either role. Again, as in San Francisco, it worked only when Mrs. Adler played the child's part.

---

*Oakland Public Schools information indicates that this neighborhood is predominantly Black (approximately 65%) with some white (approximately 25%) and the remainder Spanish Surname and Asian. All schools in the area qualify for Title I ESEA monies.

Mrs. Adler skipped coverage of the Responsive Program Principle, so I was unable to determine if this part of the session worked. All four parents were interested in the toys, knew how to play the game, and said they intended to use them with their child.

Note: I have called all nine parents who attended at the two sites one week later. Seven of the nine had used the toys with their children.

2--Haven't had time yet.

3--Played once.

2--Played twice.

1--Played four times.

1--Played six times.

Average time played--"about ten minutes."

The following table summarizes responses to key questions in the follow-up interview with the nine parents. (Please refer to Table 1, page 7.)

Conclusions

1. Test plan failed to anticipate that although this was a test of a subsequent lesson, it would be a new experience for the test subjects.
2. There were major differences in the teaching skills of the librarians recruited at the two sites.
3. Both librarians reacted positively to materials. Both liked the filmstrip, but both thought it could be longer.

## TABLE 1

## RESPONSES TO KEY QUESTIONS (Follow-Up Interview)

| | San Francisco | Oakland |
|---|---|---|
| Number of Parents | 5 | 4 |
| **Value of Session.** | | |
| (a) Very useful--quite worth attending. | 2 | 0 |
| (b) Useful. | 3 | 2 |
| (c) Not so useful--Could have done something else with my time. | 0 | 2 |
| **Did librarian help to learn how to use the toy?** | | |
| (a) Very helpful. | 3 | 0 |
| (b) Helpful. | 2 | 3 |
| (c) Not very helpful. | 0 | 1 |
| **Did the filmstrip help to learn how to use the toy?** | | |
| (a) Very helpful. | 3 | 3 |
| (b) Helpful. | 1 | 0 |
| (c) Not very helpful. | 1 | 1 |
| **Would you attend another session to learn how to use a new toy?** | | |
| (a) Yes | 4 | 3 |
| (b) No | 1 | 1 |

4. Role-playing did not work well on either site.
5. The Responsive Program Principle was presented at only one site, and there the results were disappointing.
6. All the parents learned how to use the toy and how to play the game.
7. The majority of the parents report that they have since played the game with their children. Seven of the nine parents said they would attend a second session.
8. Although it is hard to support on the basis of only three tests (these two and the prototype test), it appears that the present model of the lesson is adequate only in the hands of a relatively experienced teacher (e.g., Mrs. Lopez). The present form of the lesson was not adequate for Mrs. Adler, which suggests one of four things:
    (a) The problem is peculiar to Mrs. Adler and the Oakland test site. (Not a valid problem.)
    (b) "Librarians" should be selected on the basis of previous teaching experience. (Selection problem.)
    (c) If inexperienced librarians must be used, they should be trained. (Training problem and we aren't funded to develop training except for a self-instructional package.)
    (d) If inexperienced librarians must be used and we can't provide training, other than by a self-instructional package, we need to strengthen the model--e.g., more examples, longer description of principles, more concrete suggestions for activities, question-and-answer sections, etc.
9. This lesson and the prototype seem to be moving in the right direction, but we haven't developed a model (yet) that fully meets our design requirements.

## TASK 4: OUTLINING A SCRIPT

C

Please do not read this task until you have completed Task 3 and an evaluation form for Task 3.

Time allowed (including reading) - 1 hour

# MEMORANDUM

TO: Developers of the Librarian's Manual on Feely Bag

FROM: Doris Shin, Team Leader

SUBJECT: Producing a Sound Filmstrip to Model Role Playing.

After reviewing the Summary of the Tryout of the Feely Bag Resource Kit, which you reviewed in suggesting revisions for the Manual, it is apparent that the role-playing did not work well at either site. I think we need to produce an audio-visual aid depicting role-playing, with parents playing both the parent's and the child's roles. Undoubtedly a color movie would be effective, but it's beyond our development budget and would also make the price of the Toy Librarian's materials too high. I think we could afford a sound filmstrip if it's kept quite short. The challenge will be to produce something that carries the necessary action in both the audio and the visual components. Perhaps if we had a rough script it would give our staff something tangible to consider.

As you know, I've been working on the other AV production, but I don't have time to work on this. Please see what you can do about drafting an outline of a script. Use the Feely Bag as the toy. You'll probably want to start with a parent training session in which two parents are assigned roles of "parent" and "child" with the "parent" teaching the "child" the game, and then have the parents reverse roles. You might want to use the Feely Bag Script attached to the end of the prototype Librarian's Manual, which I gave you before we were preparing for the tryout, as a source of ideas. (See Task 2 attachment, last 2 pp.) Remember the purpose of this filmstrip is to help parents in role-playing so that they and their children can get more out of the Lending Library toys.

I don't need a complete shooting script, but I do need some concrete ideas from you about what the filmstrip might look like. Please be sure to include your ideas about the visual content that should accompany the audio. The script you're outlining should be under five minutes in duration (and probably not more than 50 film frames). I know this is a tough assignment and a very short notice. But see what you can do.

### *TASK 4: OUTLINING A SCRIPT*

*This is a difficult task since you will be given only 1 hour to work on it. Your challenge is to use the time as productively as you can. A complete shooting script is not required. Do try to outline the entire content of the sound filmstrip and to produce some detail (both audio and visual) for the first several frames.*

* * *

REMINDER. When you have finished Task 4 please fill out a copy of the DD&E Simulation Task Evaluation Form (goldenrod), then complete the General Evaluation Form (blue). When you have done this, please check to be sure you have completed the entire Battery (Biographic Form, Self-Ratings of Performance, Self-Ratings of Knowledge, Knowledge Test, 4 Simulation tasks and Simulation Evaluation Forms), and then send the completed Battery with your signed contract in the return envelope provided.

## DD&E SIMULATION TASK EVALUATION FORM

Task No.________ ID No.________

1. How long did you take to complete this task?________________________

2. How satisfied were you with the document(s) you produced? (Check one.)

☐ Highly Satisfied ☐ Satisfied ☐ Undecided

☐ Dissatisfied ☐ Highly Dissatisfied

3. Have you ever done a task like this before? ☐ No ☐ Yes

If yes, how many times?____________________

When was the most recent time (date or approximately how long ago)?

________________________________________________

4. Have you ever received any training (courses or on-the-job) in how to perform a task such as this?

☐ No ☐ Yes

If yes, please briefly describe when and where.

________________________________________________

________________________________________________

5. What modifications would you recommend to improve this simulation task?

________________________________________________

________________________________________________

________________________________________________

________________________________________________

________________________________________________

DD&E/CAB/4-75

1975 Far West Laboratory for Educational Research and Development

# Appendix D

## Scoring Responses on the DD&E Knowledge Test

Laird Blackwell

# SCORING RESPONSES ON THE DD&E KNOWLEDGE TEST

Although it may be difficult for experts to agree on the knowledge or competencies that are involved in educational DD&E, we have tried to design questions that can be used to assess knowledge and (especially) ability to apply knowledge which we feel would be important for most DD&E practitioners regardless of their specific content areas.

Scoring these questions will involve some judgments by the scorer regarding criteria and their attainment; in the following pages we have suggested criteria and methods of scoring that may help these judgments and may facilitate agreement among scorers. However, you might want to adopt your own criteria and scoring system to fit the particular needs or emphases of your program or situation.

For each of the ten questions on the knowledge test, we have suggested: 1) a checklist for responses or categories of responses, 2) criteria and methods for scoring the adequacy of responses, and 3) examples of responses that demonstrate application of a broad scope of relevant knowledge to the questions as a guide to determining a "scope" score.

The "response checklist" for each question lists those components of a response that we feel are important enough to be counted toward a rating of adequate. The "suggested criteria and methods for scoring" indicates which and how many of the components listed in the checklist we feel are required for a subject's response to be judged "adequate." That section also presents a method for deriving a "fraction score" for degree of adequacy. Although you may want to rely solely on a rating of "adequate" or "inadequate" for each response, we feel that a "fraction score" indicating how nearly a response approaches adequacy would also provide useful feedback to subjects. Suggestions for deriving a third score--the scope (of relevant knowledge applied to a question)--are also provided in this section. Whereas the other two score indicate how a response compares to minimum requirements for adequacy, the "scope score" is independent of these requirements and reflects the "breadth" or diversity of relevant knowledge applied to the question. For each question, some responses that might be judged to demonstrate the application of a broad range of relevant knowledge are suggested. A scoring scale is presented under the "scope score" section for question #1 as a guide for assigning the "scope score." At this point, we don't feel that standards or requirements for the score can be predetermined so we only present suggested guidelines for what will be a very subjective judgment on your part. Despite this subjectivity, we feel that the scope of relevant knowledge applied is an important aspect of competence and should be included in assessment.

Although you might want to develop your own criteria and/or scoring system, preliminary field testing indicates that our suggested scoring methods have some validity: knowledge test total scores for adequacy of response and scope of response were only moderately related to each other, but they each discriminated between students and professionals and were related to total scores for adequacy and scope on the simulation exercises.

Although you would probably want to refer to the scoring manual while scoring a subject's response to remind you of criteria and methods, it would be much more efficient and economical to use a separate summary sheet (like that on the next page) to record the scores, since your evaluation of many responses can then be recorded on one page. It is also suggested that all subjects' responses on one question are scored before proceeding to another question.

## SUMMARY RECORDING AND SCORING SHEET

### Knowledge Test #1: Description of Product

| "Subject" Number or Name | SCORES | | | CHECKLIST | | Comments |
|---|---|---|---|---|---|---|
| | Adequate=1 Inadequate=0 | Degree of Adequacy 0-10 | Scope 0-4 | Objectives | User Group | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

## QUESTION 1: PRODUCT DESCRIPTION

If you were given the task of developing the product described below, what important information would you need that has not been included?

> The "Victorian England" course will serve as introduction to art, literature, and social life in England during the Victorian period. The course will include selected readings from texts on Victorian social life and the arts, but will emphasize the visual arts and literature of the period as instruction in those areas as well as sources of information about social customs. "Victorian England," when produced and packaged, will include the following: 1) a book of readings, for each student, including works by and about residents of Victorian England, and 2) a paperback book for each student, of black-and-white photographs of Victorian painting, sculpture, and architecture.
>
> Users will be required to read the required text, but user activities will focus on analyzing and discussing reproductions of Victorian art and several of the best-known works of Victorian authors.

A. Response Checklist

_____objectives

_____user group

B. Suggested Criteria and Methods for Scoring

I. Rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive checks for "objectives" and "user group."

II. Degree of adequacy score--score one point for each of the responses on the checklist above included in the subject's response. (In order to obtain equivalent scores for each question on the knowledge test, you will probably want to convert the point scores into scores on a standard scale, e.g., one of two possible points on question 1 would be the equivalent of 5 of 10 possible points and so would receive a score of 5 on a 0-10 scale, as would two points of a possible four on question 3.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (very limited to very broad). This score is independent of the adequacy score--a subject's response could be judged broad in scope even though it did not include all the required aspects. You might want to use the following guidelines to help assign this scope score.

0 - included nothing other than responses required

1 - included one or two responses which were not required and which you judged to be acceptable but not especially imaginative or indicative of a broad scope of relevant knowledge

2 - included a great number of responses which were not required and which you judged to be acceptable but not especially imaginative or indicative of a broad scope of relevant knowledge

3 - included one or two responses which were not required and may not even have been on the checklist which you judged to be acceptable and very imaginative or indicative of a broad scope of relevant knowledge

4 - included many responses which were not required and may not even have been on the checklist which you judged to be acceptable and very imaginative or indicative of a broad scope of relevant knowledge

b) examples of possible imaginative responses indicating application of a broad range of relevant knowledge

Evaluation plan, teacher training necessary, entry level requirements, costs and other constraints, mode of instruction, interfacing with existing programs.

C. Sample Responses and Suggested Scoring

I. Adequate sample response

formal set of behavioral objectives

cost of materials

age level of subjects intended to use materials

Since this response includes both criterial items on the checklist, it would receive the maximum possible two points ("10" on the 0-10 scale) and a rating of "adequate," but would only receive a "1" on the 0-4 scope scale because one acceptable but not required component ("cost of materials") which was judged to be not especially imaginative or indicative of a broad scope of relevant knowledge was included.

II. Inadequate, but broad scope sample response

- target audience
- mode of instruction
- duration of course
- estimated price of package
- method of dissemination

Since "objectives" (crucial information for a developer) was not included, this response received only one point (a score of "5" on the 0-10 scale) and was judged inadequate. However, it received a scope score of "3" for the scope of knowledge indicated by some of its other components (especially "mode of instruction").

III. Adequate and broad scope sample response

- learner characteristics, i.e., age, ability, SES, year in school
- placement in curriculum, i.e., preceded by and followed by which courses
- preferred delivery system
- anticipated outcomes/objectives
- availability of different media
- constraints of time and money

Adequate; (score of "10" on 0-10 scale); "4" on 0-4 scope scale.

## QUESTION 2: OBJECTIVES

On the basis of the following description, specify appropriate terminal objectives for this product.

> The "Plant Life of New Zealand" film will include still photographs as well as movie sequences of New Zealand's most common plants both wild and domestic. The film will be divided into several segments, each covering a different family of plants. Each segment will be followed by a short quiz on plant identification and associated knowledge. The film is intended to familiarize the high school biology or ecology student with New Zealand's common plants and their families and to teach the student what plants are edible and what plans are poisonous.

A. Response Checklist

I. Form

_____behavioral (i.e., specify what the student should be able to do)

II. Specifications

_____conditions under which the behavior will be accomplished (e.g., given photographs of _ _ _ _ _)

_____criteria for scoring student performance (e.g., 80% of plants correctly identified)

III. Content

_____identify plant families or members of families

_____distinguish between or identify poisonous and edible plants

B. Suggested Criteria and Methods of Scoring

I. Rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive one check for each of the three categories (i.e., "form," "specification," and "content").

II. Degree of adequacy score--score one point for each check with a maximum of one counted for each of "specifications" and "content." (Convert point score into equivalent on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (from very limited to very broad). You might want to use the guidelines suggested under question #1.

b) examples of possible imaginative responses indicating application of a broad range of relevant knowledge

Give characteristics which distinguish plant families or poisonous and edible plants, uses of plants, repeat quiz until achieve 100% identification level.

C. Sample Responses and Suggested Scoring

I. Inadequate sample response

> students will be able to identify and distinguish edible and poisonous plants common to New Zealand
>
> students will be able to classify plants common to New Zealand according to family

Since there are no conditions (given pictures of plants, real plants, one plant at a time, names of plants) and no criteria for determining attainment of objectives, this response received "2" of "3" required points (a score of "7" on the 0-10 scale) and was judged inadequate. It received a "0" for scope.

II. Adequate and broad scope response

> students will be able to:
>
> recognize and identify by name and family the common New Zealand plants from their pictures
>
> summarize the major characteristics of the common New Zealand plants given their names
>
> identify the common New Zealand plants given their major characteristics
>
> describe the environment in which various types of common New Zealand plants grow
>
> distinguish between edible and poisonous common New Zealand plants upon encountering them
>
> attainment of the above objectives are subject to a pre-selected criteria of satisfactory performance

Adequate; ("10" on 0-10 scale) since has form, specifications, and content; "4" on scope for objectives about environment, characteristics, and identification given characteristics.

## QUESTION 3: ASPECTS OF TRYOUT

An instructional film on the behavior and attitudes a woman might adopt during pregnancy to help her prepare for child-birth has been developed by a community health service to be used in conjunction with reading, discussion, and practice in a "store-front" weekly course for pregnant women. In order to try out this film on as many people with as few administrative problems as possible, the developers plan to show it to several hygiene classes in a nearby high school and then test for knowledge as well as interview for attitudes toward the film.

On the basis of this description, discuss briefly at least four aspects of this proposed try-out that could affect the relevance and importance of the information it would yield.

A. Response Checklist

I. User differences

_____motivation (e.g., pregnant/non-pregnant)

_____education

_____attitudes

_____age/sex/race/SES

_____other

II. Setting differences

_____storefront vs. classroom

III. Resource differences

_____equipment/comfort

IV. Use differences

_____not as part of on-going weekly course

_____not in conjunction with reading, discussion, practice

V. Method of assessment differences

_____quiz and interview rather than observation of behavior

B. Suggested Criteria and Methods for Scoring

I. Rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive at least 4 checks which must include one check for "user differences." A maximum of two checks can be counted for any one category.

II. Degree of adequacy score--score one point for each of the checks counted toward the total of four required for an "adequate" rating. (Convert to score on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (from very limited to very broad). You might want to use the guidelines suggested under question #1.

b) examples of possible imaginative responses indicating application of a broad range of relevant knowledge

Captive vs. volunteer audience, resentment of parents of high school students, test-taking "set" of students, sex of teacher, authority-based relationship of teacher and student vs. service-based relationship of health personnel and "clients."

C. Sample Responses and Suggested Scoring

I. Inadequate sample response

> discrepancies between the intelligence, age, knowledge level of the high school sample and the storefront intended users
>
> most high school students will not be pregnant
>
> the class views the film but has no reading, discussion, or practice

Since this response indicates four potentially important user differences and one "use" difference, it was judged too narrow to be adequate (only two points maximum can be counted for any one category) and received 3 points of 4 possible ("8" on 0-10 scale) for degree of adequacy. It received a "2" on the 0-4 scope scale for its two additional user differences.

II. Adequate sample response

> a) most hygiene class students will not be pregnant and so will have a different attitude toward the film than would pregnant women
>
> b) hygiene classes are mandatory whereas pregnant women are in storefront course from choice
>
> c) storefront courses include reading, discussion, and practice which hygiene class does not
>
> d) try-out is not really testing for attitudes or behaviors

This response includes a user difference (a), a setting difference (b), a use difference (c), and an assessment difference (d) and so was judged adequate ("10" on 0-10 scale). It received a scope score of "3" primarily for (b) and (d).

## QUESTION 4: JUDGING REVISIONS

List three of the basic criteria for judging the adequacy of revisions which have been made in a product.

A. Response Checklist

_____effectiveness (e.g., problem solved, greater user satisfaction)

_____compatibility (i.e., degree of fit with other parts of product)

_____impact (i.e., do other parts of the product have to be revised because of the revision)

_____feasibility

_____are there better alternative revisions

_____based on research or evaluation data

B. Suggested Criteria and Methods of Scoring

I. Rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive three checks total whick must include one, and may include two, for "effectiveness." Two checks may be counted from any category.

II. Degree of adequacy score--score one point for each of the checks counted toward the total of three required for an "adequate" response. (Convert to score on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (from very limited to very broad). You might want to use the guidelines suggested under question #1.

b) examples of possible imaginative responses indicating application of a broad range of relevant knowledge

Is there evidence that improvements are really due to the revisions, are revisions based on the field test results?

C. Sample Responses and Suggested Scoring

I. Inadequate sample response

do people use the product better

do they like it better

do they meet the performance objectives better

All the components of this response involve "effectiveness" criteria, so it was too narrow to be adequate and received two of the three required points ("7" on 0-10 scale). It was scored "0" on scope.

II. Adequate and broad scope sample response

a) Do revisions solve problem which initially motivated making revisions?

b) What is the evidence for determining the answers to the first question?

c) Is the overall product consistent internally (relation and balance between parts)?

d) Is the overall product consistent externally (i.e., does it conform to the objectives originally set for it)?

e) Are revisions identified for easy reference so they can be compared to the original versions?

This response has components involving effectiveness (a) and compatibility (c and d) and so was judged adequate ("10" on 0-10 scale). It also received a score of "3" for scope.

## QUESTION 5: FILMSTRIPS VS. MOVIES

Briefly compare sound movies to sound filmstrips as instructional media from the viewpoint of the developer, the distributor, and the user; list some advantages and disadvantages of each.

A. Response Checklist

I. Feasibility

____expense (i.e., movies more expensive)

____cost and availability of equipment needed to produce (i.e., movies require more)

____availability of equipment needed to use (i.e., filmstrip equipment is less available)

II. Impact

____movies more effective if dynamic action required

____movies more effective if facial expressions important

____filmstrips usually as effective as movies except when dynamic action or facial expressions important

____filmstrips are more easily edited

____filmstrips can be more easily stopped and adjusted to pace of students, they provide more opportunity for teacher input

III. Inclusiveness

Advantages and disadvantages of movies and filmstrips listed from viewpoint of:

____user

____developer

____distributor

B. Suggested Criteria and Methods for Scoring

I. Rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive at least one check for each of "feasibility" and "impact" and all three lines checked for "inclusiveness."

II. Degree of adequacy score--score one point for each of the checks counted toward the total of five required for an "adequate" rating. (Convert to score on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (from very limited to very broad). You might want to use the guidelines suggested under question #1.

b) examples of possible imaginative responses indicating application of a broad range of relevant knowledge

Movies listed in more catalogues, filmstrip package may become separated, filmstrips more individually tailored/can be stopped and edited more easily.

C. Sample Responses and Suggested Scoring

I. Inadequate but broad scope sample response

> filmstrips as compared to movies: are cheaper to produce, require less technical skill and planning to produce, are adaptable to a wider variety of topics and situations, are more easily maintained and shipped, are less economical to store and retrieve for distribution, require fewer special skills to use

This response received a "3" on the scope score (primarily for the statements about adaptability and retrieval), but received only four points of the five required for a rating of "adequate" ("8" on 0-10 scale), since no mention was made of the important differences in impact.

II. Adequate but limited scope sample response

> movies as compared to filmstrips are more expensive to produce, distribute, and use but are usually much more exciting and motivating for students especially when action is emphasized

This response was judged "adequate" ("10" on 0-10 scale) since feasibility and impact were mentioned and differences for producers, distributors, and users were included; however, it received a "0" on the scope scale since this response did not demonstrate any knowledge other than the minimum required.

## QUESTION 6: USER SATISFACTION

A programmed text has been developed to teach high-school students the rudiments of statistical inference. Indicate three questions a developer might want to answer in a tryout about user (i.e., student) satisfaction.

A. Response Checklist

_____interest in topic

_____motivation for further study or use

_____perceived worth of content

_____perceived clarity, ease of use of text

_____perceived ability to achieve objectives

_____attitudes toward mode of instruction

_____recommendations for improvement

B. Suggested Criteria and Methods for Scoring

I. Rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive a total of three checks; a maximum of two checks may be counted for any one category.

II. Degree of adequacy score--score one point for each of the checks counted toward the total of three required for an "adequate" rating. (Convert to score on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (from very limited to very broad). You might want to use the guidelines suggested under question #1.

b) examples of possible imaginative responses indicating application of a broad range of relevant knowledge

Would you recommend the text to a friend, are you happy with the pacing, was the format a help or hinderance in your learning, could you see how the material could be relevant to your daily life?

C. Sample Responses and Suggested Scoring

I. Inadequate sample response

| |
|---|
| a) can they solve statistical problems |
| b) can they state estimates with the appropriate probability statement |
| c) do they want to know more about statistical inference |

Only one of the components of this response (c) is a "user satisfaction" question, the others are questions about effectiveness. The response receives only one of the three points required ("3" on the 0-10 scale) and receives a "0" on scope.

II. Adequate and broad scope sample response

| |
|---|
| a) was the text organized in a logical and interesting way |
| b) were the examples provided helpful in understanding the material |
| c) did the text relate the subject matter adequately to possible uses in the subject's life |

This response covers "clarity/ease of use" and "perceived worth of content" and is rated adequate ("10" on 0-10 scale). It also receives a "3" on scope primarily for "c."

## QUESTION 7: REVIEW VS. FIELD TEST

What information could you obtain from a consultant review of a product that you would not expect to obtain in a field test of the product?

A. Response Checklist

_____content validity

_____appropriateness/importance of goals

_____editorial comments

_____organization/sequencing of content

_____possible additional resources

_____media compatibility with content

_____comparison to similar material on market/marketability

_____theoretical or research base

_____accuracy of content

B. Suggested Criteria and Methods for Scoring

I. Rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must have any two of the lines checked.

II. Degree of adequacy score--score one point for each line checked--maximum score is "two." (Convert to score on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (from very limited to very creative). You might want to use the guidelines suggested under question #1.

b) examples of possible imaginative responses indicating application of a broad range of relevant knowledge

Relationship of product and methodology to research, underlying theoretical assumptions, potential markets, findings of related field tests.

C. Sample Responses and Suggested Scoring

I. Inadequate sample response

| |
|---|
| review is based on logic and reasoning rather than on observation |
| review is based on the reviewer's unique background and expertise |

This very general response indicating the difference between review and field test does not really specify differences in information likely to be obtained, and so would receive no points for adequacy ("0" on 0-10 scale) or scope.

II. Adequate but limited scope sample response

| |
|---|
| a) whether the content is accurate or not |
| b) whether the important points are being stressed |

This response addresses accuracy (a) and content validity (b) so would be judged adequate ("10" on a 0-10 scale), but would be scored "0" for scope.

III. Adequate and broad scope response

| |
|---|
| a) adequacy and correctness of content |
| b) suggestions for alternate ways of presenting material |
| c) critical competitors |
| d) liklihood of adoption by target audience |
| e) appropriateness of cost |
| f) other appropriate audiences |
| g) adequacy of objectives |

This response is adequate ("10" on 0-10 scale) and would receive a "4" for scope primarily for b, c, d, e, and f.

## QUESTION 8: QUESTIONS FOR REVIEWERS

A large oil company has asked your R&D agency to develop a method of getting information about the uses and processing of petroleum and the present research taking place in the industry to upper elementary schools. A decision was made to develop a self-contained presentation in the form of a 16mm, 45 minute colored film. The film, shot on location, explains how petroleum is refined, why new methods of refinement are needed and what is being done to find new methods. In addition to the film, a short booklet for each student has been developed which shows, in a cartoon sequence, the many uses petroleum has in our society.

Although you have been working with the content expert provided by the company, no one else outside your agency has seen your product. Below are three people who have agreed to review your product. For each reviewer, list at least three questions you would ask them about your product.

1. A 5th grade teacher
2. A sixth grade student
3. The Oil Company Representative who initially asked you to produce the product.

A. Response Checklist

I. 5th Grade Teacher

_____a) instructional effectiveness (e.g., will students understand it, what will they get out of it, is language level appropriate)

_____b) user satisfaction (e.g., will students enjoy it, is it too long)

_____c) practicality of use (e.g., would you use it, can you run film, can you incorporate it into your class)

_____d) recommendations for revisions

II. 6th Grade Student

_____a) instructional effectiveness (e.g., what do you think you learned, knowledge questions about use/processing of oil, language level appropriateness)

_____b) user satisfaction (e.g., did you like it, would you rather learn some other way, would you like to find out more)

_____c) practicality of use (e.g., do you need additional material to use it)

_____d) recommendations for revisions

III. Oil company representative

_____a) instructional effectiveness (e.g., do you think the students will get what you want them to out of this)

_____b) user satisfaction (e.g., does it say what you want it to)

_____c) practicality of use (e.g., can you afford it in its present form, can you get schools to use it)

_____d) recommendations for revisions

B. Suggested Criteria and Methods for Scoring

I. Rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive three checks for each of the three people to be questioned. For each person, two checks from any category may be counted toward the required three.

II. Degree of adequacy score--score one point for each of the checks counted toward the total of nine required for an "adequate" rating. (Convert to score on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (from very limited to very broad). You might want to use the guidelines suggested under question #1.

b) examples of possible imaginative responses indicating application of a broad range of relevant knowledge

Promotability, what kind of evidence will convince you that the product is good, what are your reactions to the teacher and student comments, did you regard the film as propaganda, what would you like to learn.

## C. Sample Responses and Suggested Scoring

### I. Inadequate but broad scope response

Teacher

a) can you integrate this film into your curriculum

b) do students talk about the film among themselves

c) do the students seem able to make the transition from the film to reality

Student

d) did you enjoy the film

e) explain why new methods are used to refine petroleum

f) what would you like to learn about petroleum that was not covered in the film or booklet

Oil Company Representative

g) is the film accurate

h) what kind of general impression does the film create in elementary school children

i) is this product what your company had in mind

The parts of the response dealing with the teacher and the student are adequate and broad in scope, they cover effectiveness (c and e), user satisfaction (b and d), feasibility (a), and recommendations (f) and would rate a "3 or 4" on scope primarily for b, c, and f. However, only "i" is an acceptable question for the oil company representative, since "g" would be known already from the consultation with the oil company's content expert and "h" should be asked of teachers or students. The overall response, then, would receive 7 of the required 9 points ("8" on the 0-10 scale) and would receive a "3" or "4" for scope.

## QUESTION 9: DIFFICULTIES IN USE

If the subjects you have selected for the tryout of your product have difficulty in using the product effectively, what factors other than the product itself could have caused the difficulty?

A. Response Checklist

_____users (e.g., lack of motivation, inadequate ability)

_____packaging (e.g., rough, unattractive)

_____setting/facilities/time allotted

_____interfacing (e.g., not in same context or sequence as will be used)

_____inadequate directions/insufficient background information

_____evaluation procedures

B. Suggested Criteria and Methods for Scoring

I. Rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive three checks--a maximum of two may be counted for any category.

II. Degree of adequacy score--score one point for each of the checks counted toward the total of three required for an "adequate" rating. (Convert to score on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (from very limited to very broad). You might want to use the guidelines suggested under question #1.

b) examples of possible imaginative responses indicating application of a broad range of relevant knowledge

Negative attitude of conductor of tryout, inappropriate evaluation criteria, lack of safeguards to prevent improper use, group interaction problems, resistance to trying something new.

C. Sample Responses and Suggested Scoring

I. Inadequate sample response

a) using the wrong medium to provide the instructions (e.g., written instructions to elementary school children

b) giving the wrong directions

c) poor administration of tryout

This response describes three factors related to directions and instructions and so only receives two points of the three required for a rating of adequate ("7" on 0-10 scale). It might receive a "3" on scope for "a."

II. Adequate and creative sample response

conditions of test: a) teacher enthusiasm, b) teacher preparation,
c) distractions, d) time alterations,
e) variations in instructions

f) subject characteristics

g) subject and test condition interactions

This response covers users, setting, and direction factors and would be judged adequate ("10" on 0-10 scale). It also would receive a "3" for scope primarily for a and g.

## QUESTION 10: STEPS IN PRODUCING A FILMSTRIP

Outline the sequence of major steps (at least five but not more than ten) that are most important in the production of a sound filmstrip between (a) the definition of the instructional objectives and (b) the collection of the first validation (pilot test) data.

A. Response Checklist

I. Outline

_____treatment (i.e., rough outline of what should be in film sequencing)

II. First version

_____story board

_____rough script

III. Preparation for Evaluation

_____development of pilot test instrument, selection of target group for evaluation

IV. Preliminary production

_____(arrange for) production of art work

_____(arrange for) review of plan and art work

V. Final production

_____production/integration of slide sequence and audio tape

B. Suggested Criteria and Methods for Scoring

I. Rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive at least five checks including at least one for each category.

II. Degree of adequacy score--score one point for each category which has a check. (Convert to score on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (from very limited to very broad). You might want to use the guidelines suggested under question #1.

b) examples of possible imaginative responses indicating application of a broad range of relevant knowledge

Relate instructional steps to objectives, synchronize the audio and the visuals, determine the constraints on film, consults with specialists.

## C. Sample Responses and Suggested Scoring

### I. Inadequate sample response

a) select site for filming

b) prepare personnel at site

c) make the filmstrip

d) try it out and revise as necessary

e) pilot test it

This response is too general to be an adequate description of the essential steps in the production of a filmstrip; "make the filmstrip" is covered in one step. This response would probably receive no points for adequacy unless "c" was counted as one of the steps in which case it would receive 1 point ("2" on 0-10 scale). It would probably receive a "0" or "1" for scope.

### II. Adequate sample response

a) prepare criterion items

b) prepare a product design

c) draft a storyboard and script for audio

d) review by subject matter expert

e) create original set work required

f) audition for soundtrack

g) create master copy of video and audio; synchronize and release

All the major steps are covered in this response and so would be rated adequate ("10" on the 0-10 scale); it should be noted that a response doesn't have to include the "shop" vocabulary (e.g., "storyboard") in order to be judged adequate. This response would probably receive a "3" for scope primarily because of a and d.

# Appendix E

# Scoring Manual for DD&E Simulation Tasks

Laird Blackwell

## SCORING MANUAL FOR DD&E SIMULATION TASKS

Although the simulation tasks will probably be more difficult and time consuming to score than the knowledge test, we have tried to provide suggestions as to criteria and methods of scoring these tasks similar to those we presented for the knowledge test. You may want to develop your own criteria and/or scoring system to reflect the task and performance level requirements of the position or job that the test taker might be expected to perform.

As with the scoring guidelines for the knowledge test, we have suggested for each simulation task: 1) a checklist for responses or categories of responses, 2) criteria and methods for scoring the adequacy of responses, and 3) examples of responses that demonstrate application of a broad scope of relevant knowledge to the questions as a guide to determining a "scope" score. In addition we have provided examples of how selected protocols from the preliminary field test might be scored using our guidelines.

The "response checklist" for each question lists those components of a response that we feel are important enough to be counted toward a rating of adequate. The "suggested criteria and methods for scoring" indicates which and how many of the components listed in the checklist we feel are required for a subject's response to be judged "adequate." That section also presents a method for deriving a "fraction score" for degree of adequacy. Although you may want to rely solely on a rating of "adequate" or "inadequate" for each response, we feel that a "fraction score" indicating how nearly a response approaches adequacy would also provide useful feedback to subjects. Suggestions for deriving a third score--the scope (of relevant knowledge applied to a question)--are also provided in this section. Whereas the other two scores indicate how a response compares to minimum requirements for adequacy, the "scope score" is independent of these requirements and reflects the "breadth" or diversity of relevant knowledge applied to the question. For each question, some responses that might be judged to demonstrate the application of a broad range of relevant knowledge are suggested. A scoring scale is presented under the "scope score" section for question #1 as a guide for assigning the "scope score." At this point, we don't feel that standards or requirements for the score can be predetermined so we only present suggested guidelines for what will be a very subjective judgment on your part. Despite this subjectivity, we feel that the scope of relevant knowledge applied is an important aspect of competence and should be included in assessment.

Although you might want to develop your own criteria and/or scoring system, preliminary field testing indicates that our suggested scoring methods have some validity: total scores on simulation tasks for adequacy of response and scope of response were only moderately related to each other, but they each discriminated between students and professionals and were related to total scores for adequacy and scope on the knowledge tests.

Although you would probably want to refer to the scoring manual while scoring a subject's response to remind you of criteria and methods, it would be much more efficient and economical to use a separate summary sheet (like that on the next page) to record the scores, since your evaluation of many responses can then be recorded on one page. It is also suggested that all subjects' responses on one question are scored before proceeding to another question.

SUMMARY RECORDING AND SCORING SHEET

SIMULATION TASK #1:

Preparing Guidelines for Developers of the Librarian's Manual

|  | SCORES |  |  | CHECKLIST |  |  |  |  |  |  |  |  |  |  |  |  |  |  | SCORE 1-4 |  |  | COMMENTS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| "Subject" Number or Name | Adequate=1 Inadequate=0 | Degree of Adequacy 0-10 | Scope 0-4 | Objectives |  | Consistent with Information |  |  |  |  | Instructions |  |  |  |  | Directions |  |  | Writing |  |  |  |
|  |  |  |  | behavioral | content | purpose | format | user group | user activities | constraints | intro. parents | filmstrip | role-playing | answers questions | others | detail | not actual manual | length | organization | style | language |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

## TASK 1: PREPARING GUIDELINES FOR DEVELOPERS OF THE LIBRARIAN'S MANUAL

This task calls for the preparation of a memo providing guidelines for the developers of Section 2 of the Librarian's Manual for the parent training session on the Feely Bag toy. You will have 1 1/4 hours to complete it. It should be no longer than the equivalent of three (3) double-spaced typed pages (900 words). You have been provided with several documents (e.g., Feely Bag toy description and Section 1 of the Librarian's Manual--both attached to this memo, a description of the Toy Lending Library, a description of your role as a member of the development team) which should provide you with the information you will need to accomplish this task. You are free to re-read, refer to, and use these documents in any way during this part and later parts of the simulation. However, please do not refer to documents included in later tasks.

Remember, you are not responsible for actually developing the Librarian's Manual. Rather, you need to extract the relevant information from the documents provided and organize it in a memo so that the development team will know for whom the manual is to be written, in what form it is to be written, what it is supposed to accomplish, etc. Your memo should also outline the content of Section 2 of the manual, i.e., the steps the Librarian should follow in the parent training session (e.g., introducing parents, leading role-playing, etc.)

---

A. Response Checklist

(Check those responses indicated below that are included by the subject. Leave a response line blank if, by your judgment, the subject's response is not sufficiently similar to or does not meet the conditions specified in the designated response.)

I. The elements of the description of the manual (i.e., its specifications) are consistent with the information in the provided documents.

_____a) purpose--to enable the Librarian to plan and conduct the parent training session
e.g., to provide all the information needed for the leader of a training session; to familiarize the leader with the toy and its purposes and to help the leader with techniques and methods for similarly familiarizing parents

b) objectives

_____1. are stated in behavioral terms (i.e., specify what the user should be able to do)

2. Librarian should be able to:

_____ plan the session

_____ introduce the toy and its purposes

_____ organize the role-playing session

_____ review the Responsive Program Principles

_____ answer questions asked by the parents

c) format

_____1. instructions to Librarian should be double-spaced, numbered, and listed in order as they are to be used

_____2. sections to be read aloud to parents should be triple-spaced and boxed

_____ d) user group (i.e., Librarians)--adults who are part of the community of the target parents
e.g., adults who live in the same community as the parents; "neighbors" of the target parents who have about the same incomes as those parents

_____e) user activities--Librarian will read the Manual before conducting the training session and will refer to it during the session

_____f) constraints--appropriate difficulty
e.g., no technical terms, not above an 8th grade reading level

II. The outline of the content of the Manual includes instructions to the Librarian for his/her most important tasks.

a)

_____1. operating and discussing the filmstrip

_____2. leading role-playing

_____3. answering questions from parents

b)

_____1. planning the session

_____2. introducing parents to each other

_____3. introducing the toy and its purposes

_____4. specifying what parents need to do

_____5. discussing the Responsive principle

III. The guidelines are well-written.

Rate each of the following on a 4-point scale where:

1 - very poor/inadequate

2 - weak, of marginal adequacy

3 - good/quite adequate

4 - excellent

_____a) organization: logical sequence, smooth transitions

_____b) style: clarity, fluidity

_____c) language: grammar, spelling, appropriate use of words

(Check each line a-c for which the rating was "3" or "4". If only one rating is "2", it also receives a check; if more than one rating was "2", give only one of them a check.)

IV. Directions/instructions were followed.

_____a) provides enough detail in specifications/guidelines so that developers could develop the actual Librarian's Manual

_____b) presents guidelines for developers rather than attempting to develop the Manual

_____c) is of appropriate length (i.e., no longer than the equivalent of three double-spaced typed pages--900 words

B. Suggested Criteria and Methods for Scoring

(These are only suggested criteria and methods; you may want to develop your own.)

I. Overall rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive: a) checks on all the lines for I, III, and IV except that only two objectives (Ib2) are required, b) checks on all three lines for IIa and on two lines for IIb.

II. Degree of adequacy score--score one point for each of the checks counted toward the 20 required for an "adequate" rating. (In order to obtain equivalent scores for each of the simulation exercises, you will probably want to convert the point scores into scores on a standard scale, e.g., 4 of a possible 20 points on Task I would receive a score of "2" on a 0-10 scale as would 3 of a possible 15 points on Task 3.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (very limited to very broad). This score is independent of the adequacy score--a subject's response could be judged broad in scope even though it did not include all the required aspects. You might want to use the following guidelines to help assign this scope score.

0 - included nothing other than responses required
1 - included one or two responses which were not required and which you judged to be acceptable but not especially imaginative or indicative of a broad scope of relevant knowledge
2 - included a great number of responses which were not required and which you judged to be acceptable but not especially imaginative or indicative of a broad scope of relevant knowledge
3 - included one or two responses which were not required and may not even have been on the checklist which you judged to be acceptable and very imaginative or indicative of a broad scope of relevant knowledge
4 - included many responses which were not required and may not even have been on the checklist which you judged to be acceptable and very imaginative or indicative of a broad scope of relevant knowledge

b) examples of possible "imaginative" responses indicating application of a broad scope of relevant knowledge

Instructions to Librarian--debriefing session with suggested questions for Librarian to ask parents; peer evaluation of role-playing.

C. Sample Responses and Suggested Scoring

The following protocol, reproduced in its entirety (with comments keyed to numbered parts of the response) should provide some guidance as to possible criteria and scoring of Task #1.

---

## PROTOCOL

### TASK 1: Preparing Guidelines for Developers of the Librarian's Manual

Background Information

Background information on the Parent/Child Toy Lending Library is included in the Librarian's Manual for Session 1, and parents will already be familiar with it. Therefore, detailed specifications of the theory and pedogogical orientation underlying the Lending Library is not included in the design for Session 2. Reference to this fact will be made at an appropriate point in Manual 2. (see format section below) ①

Product Description

Manual 2 of the Parent/Child Toy Lending Library is designed to serve as the guide for the Librarian who will be leading the parent training session explaining the educational use of the Feely Bag toy, the second toy in the eight-week lending-library course. ② Users of the product practically comprise two groups: 1) the "librarian" or other service personnel in charge of administering the Parent/Child Toy Lending Library; since agencies scheduled as target audiences (buyers/renters) of the Toy Lending Library include public libraries, churches, community centers, hospitals, schools, day-care centers, or simply interested parents who organize to use the library, the Manual must be appropriate to the middle-level professional staff of all these agencies. ③ Therefore, the language and instructions must be clear, direct, not couched in heavily technical or theoretical terms. ④ 2) The ultimate users of the Parent/Child Toy Lending Library are the parents themselves. Potential users of this group include all families whose income is above the level qualifying them for Head Start but not high enough to afford private nursery school. As this criterion defines 3/5 of the country's population, users will cover a broad spectrum of social, educational and professional levels. Instructions included in the Manual must be phrased in a language comprehensible to all groups; it should reflect approximately an 8th grade reading comprehension level. ⑤

The Manual itself is to be used only by the librarian, primarily before the session (as orientation material) and also referred to during the session. ⑥ It should provide a thorough and clear description of the Feely Bag toy, of the educational purposes of the toy, and of the orientation, practice, and role-playing in which the parents will participate. ⑦ Objectives of Manual 2 are:

1. To familiarize the librarian (or other "trainer") with the Feely Bag toy and with how it is used.
2. To enable the librarian to explain this toy and its use to the parent users during a formal training session. ⑧

Activities which the librarian/course leader will engage in while working with the manual are: reading the manual; working with the Feely Bag in order to become familiar with its characteristics; viewing the accompanying filmstrip; reading the accompanying instructions to parents. Other user activities, such as practice training sessions and reading additional source material, may be added at the user's discretion.

Production, user, and distribution constraints on Manual 2 are: it is to be printed on newsprint, thereby making it cheaper to users, and making later additional printings, as cheap as possible, it is to be spiral bound between staff cardboard covers; the printer can accommodate no more than 500 of these on a single run; the light paper construction is necessitated partly by shipping costs - all materials for each session of the 8-week Parent/Child Toy Lending Library course will be delivered parcel post (U.S. mail). Users require direct, clear language unencumbered by technical jargon or complicated rationales. ⑨ Manual 2 is to have the following characteristics:

1. No more than 20 pages in length (printed both sides).
2. Printed on newsprint pages, 8 1/2 x 11", in double-spaced gothic script with 1 1/2" margins on all sides.
3. Bound between 2 glossy stiff cardboard covers: the front cover will include a) the title ("Librarian's Manual, Session 2: Feely Bag toy", b) a 3 color cartoon illustration of the toy, and c) the developer's name and the words "Parent/Child Toy Lending Library". Manual 2 will have a tan ground with a blue, red and yellow illustration. ⑩
4. First edition print of 500.

The format of Manual 2 is as follows:

1. Introduction to librarian (2 pages, no more).
   a. An introductory blurb describing the Feely Bag toy (on page 4 of Task 1 instruction).
   b. Brief discussion of its purpose and of the principle illustrated by this toy. ⑪

2. Instructions on how to demonstrate the toy (1/2 page).

3. Librarian's introduction to parents:

   a. Statement of the purpose of this introduction (1 paragraph) - something about need to introduce this toy and how its purpose differs from that of the first session's toy. ⑫

   b. Instructions to parents (to be read or referred to in discussion, at leader's discretion). Instructions to parents should be boxed, triple space, in italics. ⑬ They should refer to the following major points and elaborate on these:

      1. Description of the game: "One set of shapes are placed in the drawstring bag..." (see p.4 of Task 1 instructions).
      2. The child should be asked once a day if he or she wants to play with the toy.
      3. The child may change the rules of the game at any time and the parent should acquiesce.
      4. Parents should stop the game when the child loses interest. ⑭

4. Instructions to distribute Feely Bags to each parent, to let them examine them a while.

5. Interlude to show the filmstrip.

6. Instructions on the purpose of the role-playing session (referring to analogous section in Manual 1 for greater detail) and on how to initiate and coordinate the role-playing. Include sample protocols from pilot test as example of role-playing (interlude for role-playing). (3 pages) ⑮

7. Guidelines for debriefing session (5-6 pages) including:

   a) A list of questions to ask parents (hypothetical questions to stimulate discussion: e.g., "what would you do if your child was unable to guess any shapes the first time around? How could you make it easier to start?" "Suppose your older child wanted to play too - how could you make it fun for both of them, help the older child to teach the younger?", etc.) ⑯

b) Samples of comments made by pilot users (e.g., "great fun", "hard to use at first but very rewarding") to add human element/enthusiasm/etc. A cheery closing note.

8. "Any questions?"

---

KEYED SCORING GUIDE TO PROTOCOL FOR TASK 1

| Number on response checklist | Numbered part of subject's response |
|---|---|
| Ia: purpose | ② and ⑦<br>taken together indicate that the purpose of the manual is to familiarize the librarian with the uses of the toy and with methods of teaching these uses to parents (so Ia receives a check). |
| Ib: objectives | ⑧<br>is stated in behavioral terms (so Ib1 is checked) but is limited to enabling the librarian to explain the toy and its uses. A developer of the manual should know that the manual needs to focus specifically, on enabling the librarian to organi role-playing etc. - objectives which are specifie in the document "Your role as a member of the development team." The response receives only 1 of the required 2 checks for Ib2. |

| | |
|---|---|
| Ic: format | ⑬<br>specifies that instructions to be read to parents should be triple-spaced and boxed. (so Ic2 is checked). The entire "format" section of the response numbers and lists instructions and procedures for the librarian; assuming that this is a model for the manual developer to follow, Ic1 also receives a check. |
| Id: user group | ③ and ⑤<br>taken together imply that the librarians will be adult members of the target communities (so Id receives a check). |
| Ie: user activities | ⑥<br>indicates that the librarian will read the manual before the session and will refer to it during the session (so Ie receives a check). |
| If: constraints | ④, ⑤, and ⑨<br>specify a clear, non-technical, 8th grade reading level (so If receives a check). |
| IIa: essential instructions for librarian | ⑮ and ⑯<br>indicate that instructions for conducting role-playing and answering questions are to be included. However, since there is no specification that instructions for running and discussing the film are to be included, the response receives only 2 of the required 3 checks for IIa. |
| IIb: other instructions for librarian | ⑪, ⑫, and ⑭<br>specify that instructions for the librarian relevant to the Responsive Principle, introducing the toy and its purposes, and suggesting things for parents to do with their children while playing with the toy are included (so IIb receives the maximum scorable 2 checks). |
| III: | Style and language are rated "4" for excellence; the organization was judged to be generally good though somewhat redundant so was rated "3". (So III received 3 checks). |

IV. Checks were given for all three categories involving the following of directions; the response was judged to be detailed enough to enable development of the manual, but was still an outline of what should be included in the manual rather than an attempt to actually write the manual.

A total of 18 of the 20 checks required for an "adequate" rating were received, so this response might be judged inadequate, but would receive a "9" on the 0-10 degree of adequacy score.

The response might receive a "4" for scope of applied relevant knowledge primarily for the descriptions of: developer constraints and specifications (10), rules parents should follow or explain to their children in using the toy (14), "debriefing" procedures for the librarian to follow (16), and the relationship between sections 1 and 2 of the manual (1).

## TASK 2: PREPARING FOR THE TRYOUT OF THE RESOURCE KIT

In this task you are to: (1) describe what the Research and Evaluation Team should look for in selecting test sites, librarians, and parents, and (2) list the questions (separately for librarians, parents, and observers) that you would like to have answered in a tryout of the Librarian's Manual that we have developed.

The list of questions is the more important of the two activities. It should take no longer than the equivalent of two (2) typewritten double-spaced pages.

The description for the Research and Evaluation Team should be brief-- one or two paragraphs.

This task should take no more than 30 minutes.

---

A. Response Checklist

I. Description of what to look for in selection of tryout sites and users is consistent with information in provided documents.

_____a) site--appropriateness to intended use
e.g., in community--school, library, church, union, etc.

_____b) librarians--community members

_____c) parents--appropriate (as designated) income level, parents of 3-4 year olds

II. Questions for the librarians, parents, and observer are appropriate, useful, and comprehensive.

(Check the appropriate category for each of the subject's questions--a category may, then, have more than one check.)

a) Librarian

_____1. instructional effectiveness
e.g., were instructions clear, did it give you enough guidance, did it help you feel comfortable and competent in running session

_____2. user satisfaction
e.g., did you enjoy using the manual and running the session, would you like to do it again

_____3. practicality of use
e.g., did you have enough time to prepare, could you refer to the manual easily during the session

_____4. recommendations for revisions

b) Parents

_____1. instructional effectiveness
e.g., was the Librarian a help in learning how to use the toy, can you now explain the purposes and use of the toy

_____2. user satisfaction
e.g., did you enjoy the session, would you come to another one, was the Librarian understanding/ organized

_____3. practicality of use
e.g., was the session more trouble than it was worth

_____4. recommendations for revisions

c) Observer

_____1. instructional effectiveness
e.g., did the parents learn how to use the toy, did they understand the Responsive principle, was the language level appropriate

_____2. user satisfaction
e.g., did the Librarian/parents enjoy the session, did the parents resent the Librarian's being in charge, did they feel the session was worthwhile, did the parents use the toys with their children following the session

_____3. practicality of use
e.g., were there any equipment problems, could the Librarian use the manual smoothly

_____4. recommendations for revisions

III. The directions/instructions were followed.

_____a) questions listed separately for Librarian, parents, and observer

_____b) appropriate length (i.e., 1-2 paragraphs for site and subjects; no longer than equivalent of two double-spaced typed pages--600 words--for questions)

B. Suggested Criteria and Methods for Scoring.

I. Overall rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive: a) checks on all lines for I and III, and b) a total of three checks for each of IIa, IIb, and IIc where at least one and no more than two of those checks is for "effectiveness."

II. Degree of adequacy score-- score one point for each of the checks counted toward the 14 required for an "adequate" rating. (Convert point score into equivalent on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (very limited to very broad). This score is independent of the adequacy score--a subject's response could be judged broad in scope even though it did not include all the required aspects. You might want to use the guidelines suggested under Task #1 to help assign this scope score.

b) examples of possible "imaginative" responses indicating application of a broad range of relevant knowledge.

Consideration of feasibility (cost, distance) of sites, consideration of variety of sites/parents/librarians, specification of follow-up questions for parents after trying toy with child.

C. Sample Responses and Suggested Scoring.

The following protocol, judged to be inadequate, reproduced in its entirety (with comments about scoring keyed to numbered parts of the response) should provide some guidance as to possible criteria and scoring of task #2.

## PROTOCOL

## TASK 2: PREPARING FOR THE TRYOUT OF THE RESOURCE KIT

### Conditions

The population needed to tryout this should contain no less than two librarians. Each librarian should have about six parents as students. Individuals selected to act as librarians should be adult high school graduates. ①

The test site should consist of a large room with tables large enough to conduct role playing exercises. Electrical power outlet required. Lights with on/off switch required as tryout will probably have to be conducted in the evening. ②

### Questions

The following information should be solicited:

a. Librarian

1. Were you successful?
2. Did you enjoy the activity?
3. Would you conduct a similar session?
4. How long did it take you to prepare?
5. What additional aids could you use to make your job as librarian easier?
6. Specify problems you had in using the manual. ③

b. Parents

1. Can you use the Feely Bag toy after going through the session?
2. Which activity helped you most?
    - a. Introduction
    - b. Demonstration
    - c. Role-playing
    - d. Individual help after role-playing

3. Would you borrow and use this toy with your child? ④

c. Observer

1. Was the librarian able to lead the parents through the learning experiences with ease?
2. Note points of confusion observed during the presentation by the librarian and parents.
3. Evaluate the attitude of ⑤ the parents at each stage of the session on one to five scale.

---

## KEYED SCORING GUIDE TO PROTOCOL FOR TASK 2

| Number on response checklist | Numbered part of subject's response |
|---|---|
| Ia: | ②<br>describes physical characteristics of the room relevant to its suitability for conducting role-playing, etc., but since it makes no mention of the more fundamental criteria for test-sites involving location (e.g., "in surroundings familiar to users, easy to get to") no check for "site" would be given. |
| Ib: | ①<br>specifies only that librarians should be adult high school graduates. Since it does not specify that they should be members of the community who have about the same income level as users, no check for "librarians" would be given. |
| Ic: | To receive a check for "parents", some mention should have been made of their income level and the age of their children. |

| Number on response checklist | Numbered part of subject's response |
|---|---|
| IIa: | ③ |

The questions for the librarian cover "effectiveness" (1,6), "user satisfaction" (2,3), "practicality of use" (4), and "recommendations" (5), but some of the questions (e.g., 1 and 2) are very general and of questionable use to the field tester. You might decide they are adequate, but you would undoubtedly give little credit for "scope". The following questions from another protocol would provide more useful information:

Were you able to follow the sequence of the Manual exactly?
If not, what happened that caused you to change your sequence?
Were you able to give quick, confident answers to questions asked?
If not, what were the questions you were not prepared for?
Could you follow the Manual conveniently during the session or did you have trouble finding your place?
If you had trouble, where?
Would you want to use a similar manual in future sessions?
What changes, if any, would you like to see in the manual?

| | |
|---|---|
| IIb: | ④ |

The questions for the parents involve "effectiveness" (1 and 2) and "user satisfaction" (3) and so could be judged "adequate"; however, again they are very general and the yes/no answers likely to be obtained would be of limited usefulness. The following questions from another protocol would provide more useful information:

At what points, if any, was the librarian confused or evasive?
What unanswered questions, if any, do you still have about how to play the game or what your child can learn from the game?
Did any problems arise in the conduct of the session that lessened your enjoyment of it?
What were they?

| Number on response checklist | Numbered part of subject's response |
|---|---|
| IIc: | ⑤<br>The questions for the observer involve "effectiveness" (2), "user satisfaction" (3), and "practicality" (1) and would probably yield valuable information, so would probably be judged "adequate: (i.e., receive checks). Many other questions about procedures during the session, parent participation, and outcomes for the parents would be useful. |
| III: | The directions concerning length and separate questions for librarians, parents, and observer were followed, so IIIa and IIIb would receive checks. |

Since this response did not receive any checks for I and might not have received all the required checks for IIa or IIb (because of the generality of the questions), it would be scored from 5-11 on "degree of adequacy" and would be judged "inadequate." Very little imagination or breadth of relevant applied knowledge was demonstrated, so it might receive a "1" on the scope score for the description of the requirements of the test site room (2) and the specification of a parent attitude scale for the observer to fill out (5).

## TASK 3: RECOMMENDING REVISIONS FOR THE LIBRARIAN'S MANUAL

There are two activities involved in this task. Listed in order of importance they are:

1. Recommending revisions for the Librarian's Manual. (This product should be no longer than the equivalent of three (3) double-spaced typed pages.)
2. Critiquing the tryout. (This product should be no longer than the equivalent of two (2) double-spaced typed pages.)

You have 1 hour to complete this task.

---

A. Response Checklist

I. Suggested revisions of Manual address problems indicated in the information provided (in the Tryout Report).

_____a) role-playing
e.g., additional practice, film, models, more background information for Librarian

_____b) Responsive principle
e.g., more discussion application, more chance for feedback and input from parents

_____c) Librarian following manual
e.g., more specific instructions to follow manual step-by-step, more opportunity for Librarian to paraphrase

_____d) answering questions from parents
e.g., more examples of likely questions and how to handle them

_____e) Librarian's experience and expertise
e.g., manual made more flexible to varying degrees of Librarian skill and knowledge, optional background information

_____f) need for introductory material about program and toy if not used in sequence after a previous session

II. Characteristics identified as strengths and weaknesses of tryout are reasonable and comprehensive.

(Mark a check for each strength or weakness included which you judge to be reasonable and important. Each line, then, can have several checks.)

_____a) strengths

e.g.,

variety of sites/librarians/parents

Librarian and parents knew each other

detailed descriptions of sites and users in report

observed parents as well as Librarians

observer gave impressions and suggestions as well as record of events

parents and Librarians interviewed

follow-up on parents' actual use of toys

quantified results of interviews with parents

observer adopted procedure to situation at hand

_____b) weaknesses

e.g.,

Librarians not given manual soon enough

no objective assessment of how well objective met

no indication of how parents recruited

observer imposed self into instruction in second test site

tested session out of context (no preceding session)

observer didn't probe after initial questions (e.g., Librarian wants longer film, but what additional content is wanted?)

parent interviews not done anonymously, so may get halo effect

Librarian recruited parents--may be biased sample

confusion of what being pilot tested (i.e., manual or film)

no evidence of interview with second Librarian

same sites used as for testing prototype lesson--limited sample

III. Suggested modifications for tryout address weakness in tryout procedure or inadequacies in tryout outcomes.

(Mark a check for each modification you judge is likely to remedy a weakness in the tryout.)

_____a) selection of test site(s), librarians, parents
e.g. larger more heterogeneous samples, select librarians with more knowledge or experience in teaching or working with parents

_____b) making arrangement for the tryout
e.g. get materials to librarians earlier

_____c) directions which should be given to the librarian(s) before the session
e.g. emphasize that sequence in manual should be followed, advise librarians to get responses/feedback from parents on principles and points before proceeding further

_____d) specific directions and questions for the evaluator observing the session
e.g. observer should not impose self in session, interviews should be conducted anonymously, probe in follow-up questions to parents.

IV. The directions/instructions were followed.

_____a) revisions of manual were no longer than the equivalent of three double-spaced typed pages--900 words

_____b) the critique of the tryout should be no longer than the equivalent of two double-spaced typed pages--600 words

B. Suggested Criteria and Methods for Scoring.

I. Overall rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive a) at least four checks on I, b) at least three checks each on IIa and IIb, c) checks on at least three of the four categories for III and d) checks on both lines of IV.

II. Degree of adequacy score-- score one point for each of the checks counted toward the 15 required for an "adequate" rating. (Convert point score to equivalent on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (very limited to very broad). This score is independent of the adequacy score--a subject's response could be judged broad in scope even though it did not include all the required aspects. You might want to use the guidelines suggested under Task 1 to help assign this score.

b) examples of possible "imaginative" responses indicating a broad range of relevant knowledge.

Manual should contain brief description of teaching and group leading skills and techniques; include report of specific questions and responses in interview; need group as well as individual "debriefing."

C. Sample Responses and Suggested Scoring

The following protocol, reproduced in its entirety (with comments about scoring keyed to numbered parts of the response) should provide some guidance as to possible criteria and scoring of task #3.

---

PROTOCOL

TASK 3: RECOMMENDING REVISIONS FOR THE LIBRARIAN'S MANUAL

1. Give Librarians the manual a week ahead of the training session, so they'll have time to work out difficulties and problems.
2. Emphasize in the beginning of the manual that for those who are fairly inexperienced in leading this kind of activity, following the guide closely should be a big help until the librarian feels more confident and can branch out more spontaneously.
3. Include in more detail an explanation of the Toy Lending Library and the Far West Laboratory in the manual, which if necessary can be read to the parents.

4. Include a section for feedback from the week at home with the previous toy, right after the introduction. This is a good method of breaking the ice, getting parents to speak out informally, and encouraging them to try the toy with their children.

5. Include all the Responsive Program Principles in each manual. (It might also be a good idea to include a chart listing these principles, which can be hung up for each session.) Place special emphasis on the one being discussed at that session and give several examples of this principle in section five.

6. Give a more detailed explanation of role-playing and include one example of possible dialogue and movements which might occur with that particular toy when a child and parent use it. The librarian and a selected parent can demonstrate exactly what is in the manual for the first role-playing attempt.

7. Include a list of the most commonly asked questions of parents and possible answers.

8. Give each parent a Feely Bag Toy and a set of instructions when the librarian introduces the toy. Parents can examine the toy while the librarian is speaking about it and can read the rules on how to play the game before they view the filmstrip.

9. Make sure the manual is well-coordinated with the filmstrip. If the manual calls for role-playing or discussion of a Responsive Program Principle make sure both are previously demonstrated or discussed in the filmstrip.

10. Include a list of possible variations on the original Feely Bag Game. Include discussion and role-playing of some of these variations as part of the manual's instructions, to the librarian.

11. Rather than having the librarian review the Feely Bag Toy and it's use, let her conduct her review in the form of questions for the parents. This will let her know if parents truly understand what has been demonstrated and it will allow the librarian to clear up any confusion.

12. End the session on a pleasant and optimistic note and have the librarian encourage parents to use the toy at home.

## CRITIQUE OF THE TRYOUT

a) What are the major strengths?

1. The choice of sites was good-- both being in low-income predominantly minority neighborhoods.
2. The location of each session was normal to the participants.
3. Librarians and parents were interviewed by the observer.
4. The parents were followed up by telephone.
5. The librarians were given help in advance in order to prepare for the lesson.
6. The racial composition of the sessions was representative of the neighborhoods, although this may have been accidental since no guidelines were given to the librarians for recruitment.
7. Both librarians were members of the communities.
8. Both experienced and inexperienced teachers were used.

b) What are the weaknesses?

1. Each librarian should have been required to recruit at least 8 parents, to provide for attenuation.
2. One session was held at 10:00 a.m. when working parents would be excluded from attending. (Such parents may be more representative of those in the neighborhood.)

3. The unfamiliarity of parents with the Toy Lending Library was not anticipated.
4. The evaluator became involved in one session, so that it cannot be considered normal.
5. One librarian did not follow the manual.
6. Parents were not followed up with regard to their implementation of the Responsive Program Principles.

c) Instructions for future tryouts

1. Be sure that racial balance is representative of neighborhood.
2. Be sure fathers are included if possible.
3. Be sure that librarian is prepared to give orientation to new parents.
4. Be sure that librarian understands that he/she should follow the manual with regard to activities.
5. Be sure that evaluator does not become a teacher.
6. Have evaluator ask questions regarding implementation of the Responsive Program Principles in the follow-up telephone call: For example,
    a. Did your child make up his/her own rules?
    b. Did your child not want to play any time? If so, what did you do?
    c. Did your child want to play with the toy without using rules? If so, what did you do?

## KEYED SCORING GUIDE TO PROTOCOL FOR TASK 3

| Number on response checklist | |
|---|---|
| I: | Many of the problems indicated in the tryout report are addressed in this protocol: role-playing (a) by revision "6", the responsive principle (b) by "5", following the manual (c) by "2", answering questions (d) by "7", and introductory material (f) by "3". (The response receives 5 checks for "revisions of manual" though only 4 are counted toward the total score.) In addition, the subject has included several other revisions which show imagination and a broad understanding of potential problems and possible solutions. |
| IIa: | The listing of strengths of the tryout covers aspects of: the sites used (1 and 2), the personnel used (6 and 7), the level of experience and preparation of the librarians (5 and 8), and the scope of information-seeking by the observer (3 and 4). The characteristics listed cover a fairly broad range of aspects of tryout design and implementation and could be judged to be important strengths of the tryout. "Strengths" (IIa) would receive more than the 3 checks required to be judged "adequate." |
| IIb: | The listing of weaknesses of the tryout covers aspects of: size and bias of the sample (1 and 2), lack of appropriate context for the field test (3), non-objectivity of the observer (4), non-use of the manual (5), and lack of appropriate follow-up of parents (6). "Weaknesses" (IIb) would receive more than the 3 checks required to be judged "adequate." |
| III: | The subject's recommendations for a second tryout address three of the four areas specified in the task statement (i.e., site and personnel, preparation of librarian, directions to observer) and address most of the aspects of the tryout previously identified as weaknesses in IIb. "Modifications" (III) receives the three checks required to be "adequate." |

| Number on response checklist | |
|---|---|
| IV: | The length restrictions were not exceeded, as the two required checks for following directions were received. |

This response received all 15 of the checks required for a rating of "adequate." In addition, there were many evidences of application of broad knowledge and understanding relevant to tryouts and revisions (e.g., suggested revisions of the manual #4, #8, #9, #10, and #11; the weakness involving sample bias, #2; and the specific follow-up questions for the observer to ask the parents, #6). This response, then, would be judged "adequate" and could receive a "4" scope score.

## TASK 4: OUTLINING A SCRIPT

This is a difficult task since you will be given only 1 hour to work on it. Your task is to outline the visual and audio content of a filmstrip depicting role-playing with parents playing both the parent's and the child's roles, in a session in which a "parent" is teaching a "child" how to play with the Feely Bag. A complete shooting script is not required. Do try to outline the entire content of the sound filmstrip and to produce some detail (both audio and visual) for the first several frames.

---

A. Response Checklist

I. Outline of script is consistent with information and specifications provided.

- _____a) parents playing both roles (i.e., roles of child and parent
- _____b) includes ideas about visual as well as audio content
- _____c) explains purpose/reason for role-playing
- _____d) tries to put parents at ease/relates role-playing to them/indicates that parents may have trouble at first playing role of child but no need for embarassment
- _____e) some explanation of toy: purposes, content, or use

II. Outline of script demonstrates awareness of visual aspects of film.

- _____a) real use of visuals (i.e., visuals should convey important information even without the audio; a great deal of instruction and impact should be lost if subject's visuals were omitted)
- _____b) awareness that this is filmstrip rather than movie (i.e., series of static shots rather than motion-oriented)

III. The directions/instructions were followed.

- _____a) length: film outlined should not be over five minutes or 50 frames
- _____b) entire content outlined; enough detail (e.g., rough outline of what will be in narration and where it will come)
- _____c) first several frames in considerable detail

B. Suggested Criteria and Methods for Scoring.

I. Overall rating of "adequate" or "inadequate."

In order to receive an "adequate" rating, the subject must receive checks on all the lines in I, II, and III.

II. Degree of adequacy score-- score one point for each of the checks counted toward the 10 required for an "adequate" rating. (Convert point score into equivalent on 0-10 scale.)

III. Scope (of applied relevant knowledge) score

a) rate the subject's response as to the scope of relevant knowledge brought to bear on the question on a scale of 0-4 (from very limited to very broad). You might want to use the guidelines suggested under question #1.

b) examples of possible "imaginative" responses

Indication to parent that role-played child should not be "ideal" child, but should be "natural" and allow problems to occur; some indication that models used in film will be similar to target audience; budget considerations; detailed instructions for visuals (e.g., close-ups, pan shots).

C. Sample Responses and Suggested Scoring

The following protocol, reproduced in its entirety (with comments about scoring keyed to numbered parts of the response), should provide some guidance as to possible criteria and scoring of task #4.

PROTOCOL

TASK 4: OUTLINING A SCRIPT

| FRAME NUMBER | AVOID | VISUAL |
|---|---|---|
| 1 | You have seen a parent and child playing games with the Feely Bag toy. | Feely Bag Toy. |
| 2 | To help you learn to play games with your child, you are going to practice with the toy here. | Parent and child playing with toy. |
| 3. | This practice is called role-playing. One of you is going to pretend to be a parent and another is going to pretend to be a child. | Parent and child from frame 2 but in anotner pose. |
| 4. | Acting the role or part of a parent shouldn't be very hard, but acting the part of a child may be hard at first. | Exactly frame 3 but with an adult in place of the child. |
| 5. | Here is a man pretending to be a three-year-old child while his "mother" shows him a new game. | "Child" picking up a triangle from the table. |
| 6. | (Parent) "that shape is called a triangle. Can you find a shape in the bag which is the same as the triangle?" | Parent putting each of the 4 shapes in the bag. |
| 7. | (Parent) "see if you can find the same shape without peeking into the bag." | "Child" peeking into bag while holding triangle outside bag. |
| 8. | ("Child") "I found it." | "Child" holding up a square. |
| 9. | (Parent) "No, that is a square. Can you find this shape?" | "Parent" pointing at the triangle on the table. |
| 10. | ("Child") "here it is," (Parent) "Yes, these two shapes are the same." | "Child" holding up a triangle and parent pointing at the two triangles. |
| 11. | ("Child") "Let's find all the ones with corners." (Parent) "Alright, you find them without peeking." | "Child" holding up a square. |
| 12. | ("Child") " Is that all of them?" (Parent) "See if there are any more?" | "Child" has two triangles, two squares, and one rectangle in front of him; he is putting second rectangle on top of the other. |

| FRAME NUMBER | AVOID | VISUAL |
|---|---|---|
| 13. | ("Child") "This doesn't have corners." | "Child" holding up a circle. |
| 14. | (Parent) "Do you want to play another game?"<br>("Child") "No, I'm tired."<br>(Parent) "Alright, we'll stop now." | "Child" looks bored and is looking under the table. |
| 15. | Now you have seen how one parent pretended to be a child. He was role-playing. Did you notice how the "child's" parent let the child change the rules of the game? That is important. Also, the "child's" parent let the child stop when he was tired. | Return to frame 11. |
| 16. | Now let's see what happens when these two people trade roles. | Persons playing the parent and the child have interchanged positions. |
| 17. | ("Child") "These don't match." | "Child" is putting a square on a rectangle. |
| 18. | (Parent) "That's right. Can you find one that matches this one?" | Parent points at the square. |
| 19. | Notice that the parent is letting the child explore before starting the game. | Child is fitting the two triangles together to make a parallelogram. |
| 20-27 (approx.) | [The game is played twice, once to match circles and once to match triangles.] | [Suitable pictures] |
| 28. (approx.) | (Parent) "Can you find this shape, a rectangle, in the bag?" | Parent holding up rectangle from table. |
| 29. | ("Child") "I don't want to."<br>(Parent) "Alright, we won't play this game anymore today." | "Child" has turned away from the toy. |
|  | Notice that the parent lets the child stop when she was tired. |  |
| 30. | Now you have seen two people take turns role-playing a parent and a child. When they do this, they find out what it feels like to be in "someone else's shoes." To help you learn how to use the Feely Bag toy with your child, you are now going to get a chance to act or role play the part of a child and the part of a parent. | Return to frame 16. |

## KEYED SCORING GUIDE TO PROTOCOL FOR TASK 4

| Number on response checklist | Frame number in protocol |
|---|---|
| Ia: | 3 and 15<br>indicate that each parent has a turn playing the child and the parent, so Ia receives a check. |
| Ib: | The subject presents descriptions of the visual as well as the audio component for each frame, so receives a check for Ib. |
| Ic: | 2, 3, and 30<br>explain that role-playing the part of a child playing with the toy should help them understand problems and feelings that a child may have in such a situation. This understanding should enable them to better help their child learn how to use the toy. (Ic receives a check.) |
| Id: | 4<br>which demonstrates an awareness that playing the role of a child may be difficult at first is an important effort to relieve some of the embarrassment or tension parents might have about playing the role of a child. (Id receives a check.) |
| Ie: | As specified in the instructions for the task, the film shows role-playing in the context of a parent and child playing with the Feely Bag toy. (Ie receives a check.) |
| IIa: | IIa receives a check because visuals play a very important part in this film; sometimes the visual is an illustration of the accompanying audio, sometimes the visual presents information that is not presented in the audio (e.g. in frame #9 the parent asks "Can you find this shape?" While the visual shows her pointing to a triangle ), and sometimes the audio is a narration of what's happening in the accompanying visual (e.g., in frame #19 the child is shown fitting two triangles together while the audio has a narrator saying "Notice that the parent is letting the child explore before starting the game").<br>(In another protocol, the visuals were uninformative and contributed nothing to the film (e.g. "view of class with one person speaking," "view of librarian standing"), so IIa would not receive a check. ) |

| Number on response checklist | Frame number in protocol |
|---|---|
| IIb: | The visuals indicate an understanding of the static frame nature of filmstrips (e.g., "child pointing," "child holding up square"), so IIb receives a check. |
| IIIa: | The length restriction has not been exceeded, so IIIa receives a check. |
| IIIb and IIIc: | The entire content of the film is outlined; the first 19 frames in considerable detail and the last several in rough form (i.e. "the game is played twice once to match circles and once to match triangles --- with suitable pictures"), so IIIb and IIIc each receive a check. |

This protocol received all 10 of the checks required for an "adequate" rating. In addition, several aspects of the response (e.g. frames 14 and 29 in which the child gets bored, frames 15 and 19 which indicate that the child should be allowed to establish his/her own rules) show imagination and understanding of what would make the role-playing film most instructive. This response would be judged "adequate" (a degree of adequacy score of 10) and might receive a scope score of "3."

# Appendix F

# Miscellaneous Forms

DD&E/CAB/4-75

ID#__________

FOR FWL USE ONLY
(01)
_____03
_____05
_____07
_____08

## DD&E TRAINING PROGRAM BIOGRAPHIC DATA

Sex: F M
(circle one above)

Age______
(years)

I. EDUCATION: Highest Degree (circle one): High School, BA/BS, MA/MS, Ph.D., EdD. _____10

The following is a list of courses that are frequently mentioned as useful in educational development, dissemination and evaluation. Please mark by a number (0,1,2, etc.) the number of undergraduate, graduate or formal continuing education courses you have taken for each of the following. (Please place a zero (0) beside the category if you have taken no courses in this category.)

| | | |
|---|---|---|
| _____Curriculum Development | _____Journalism | _____11 |
| _____Educational Psychology | _____Creative Writing | _____13 |
| _____Other Psychology | _____Media Production (T.V., film, etc.) | _____15 |
| _____Other Social Sciences | _____Graphic Arts | _____17 |
| _____Teacher Education | _____Research Design | _____19 |
| _____Communication Theory | _____Test & Measures, Psychometrics | _____21 |
| _____Instructional Technology | _____Evaluation | _____23 |
| _____Marketing | _____Statistics | _____25 |
| _____Administration | _____Philosophy or Humanities | _____27 |

II. PROFESSIONAL MEMBERSHIP: Are you a member of any national professional organizations? (circle one) YES NO _____29

III. PUBLICATIONS AND PRODUCTS: Please indicate the approximate number of items for which you have been the author or co-author. (If you have never produced a specific type of item, please place a zero (0) beside that category.

| | | |
|---|---|---|
| _____Articles in professional journals | _____Tests/Assessment Instruments | _____30 |
| _____Technical R&D Reports | _____Programmed Lessons | _____34 |
| _____Evaluation Reports (of products/ programs) | _____Slide-Tapes/Audio Tapes | _____38 |
| _____Chapters in Professional Books | _____Instructional Films/T.V. | _____42 |
| _____Other Publications (Please Specify)<br>(______________________) | _____Other Products (Please Specify)<br>(______________________) | _____46 |

IV. TEACHING EXPERIENCE: Have you ever taught? YES NO (circle one) _____50

If YES, what kind of teaching? Student Substitute Regular
(circle which one(s)) _____51

How many years?__________ _____52

V. CURRENT WORK EXPERIENCE: Are you currently employed? YES No (circle one)

If YES, Is your work ____Part-Time, ____Full-Time, ____Volunteer,

____Practicum--degree requirement

Postion Title______________________________Employer____________________

Date Started_______________________Major Activities____________________

___________________________________________________________________

___________________________________________________________________

VI. PREVIOUS WORK EXPERIENCE:

Have you ever done any work (other than that described in Section V immediately above) in which you were involved in research, development, evaluation or dissemination-diffusion? YES NO (circle one)

If YES, Please complete the following, beginning with the most RECENT position.

1. Position Title______________________________From_______To_______

Employer___________________________________________________

Major Activities______________________________________________

___________________________________________________________________

___________________________________________________________________

2. Position Title______________________________From_______To_______

Employer___________________________________________________

Major Activities______________________________________________

___________________________________________________________________

___________________________________________________________________

3. Position Title______________________________From_______To_______

Employer___________________________________________________

Major Activities______________________________________________

___________________________________________________________________

___________________________________________________________________

DD&E/CAB/4-75

## FOR STUDENTS ONLY

Please check which of the following DD&E Instructional Modules you have completed.

---

PLANNING SERIES

_____ 1.0 An Overview
_____ 1.1 Problem Definition and Specification of Outcomes
_____ 1.2 Consideration of Alternatives
PLANNING/DEVELOPMENT SERIES
_____ 1.3/4.1 Introduction to Development
PLANNING/EVALUATION SERIES
_____ 1.4/5.1 Introduction to Evaluation
PLANNING/DISSEMINATION AND MARKETING SERIES
_____ 1.5/6.1 Introduction to Dissemination and Marketing

---

INFORMATION/DATA COLLECTION AND ORGANIZATION SERIES

_____ 2.1 Orientation to Collecting and Organizing DD&E Information and Data
_____ 2.2 Data Management
_____ 2.3 The Retrieval of Information Using Bibliographic Resources
_____ 2.4 The Retrieval of Information Using Special Sources

---

COMMUNICATION SKILLS SERIES

_____ 3.1 Oral Communication
_____ 3.2 Written Instructional Materials
_____ 3.3 Technical Writing: Informal Documents
_____ 3.4 Technical Writing: Formal Documents

---

DEVELOPMENT SERIES

_____ 4.1/1.3 See Planning Series
_____ 4.2 Designing Educational Products
_____ 4.3 Product Engineering
_____ 4.4 Review, Tryout, and Revision
_____ 4.5 Special Problems in Development

---

EVALUATION SERIES

_____ 5.1/1.4 See Planning Series
_____ 5.2 Measurement and Testing for Developers and Evaluators
_____ 5.3 Desing of Evaluation Instruments
_____ 5.4 Planning and Implementing Evaluation
_____ 5.5 Evaluation Problems

---

DISSEMINATION AND MARKETING SERIES

_____ 6.1/1.5 See Planning Series
_____ 6.2 Marketing Educational Products

---